DOCUMENT RESUME

ED 478 791 PS 031 433

| | |
|---|---|
| AUTHOR | Zill, Nicholas; Resnick, Gary; Kim, Kwang; O'Donnell, Kevin; Sorongon, Alberto; McKey, Ruth Hubbell; Pai-Samant, Shefali; Clark, Cheryl; O'Brien, Robert; D'Elio, Mary Ann |
| TITLE | Head Start FACES 2000: A Whole-Child Perspective on Program Performance. Fourth Progress Report. |
| INSTITUTION | Administration for Children, Youth, and Families (DHHS), Washington, DC. Head Start Bureau.; Administration for Children, Youth, and Families (DHHS), Washington, DC. Office of Research and Evaluation. |
| PUB DATE | 2003-05-00 |
| NOTE | 227p.; For Third Progress Report, see ED 453 969. |
| CONTRACT | HHS-105-96-1912 |
| AVAILABLE FROM | Head Start Information and Publication Center, Order Fulfillment Department, P.O. Box 26417, Alexandria, VA 22313-0417. Tel: 703-683-2878; Fax: 703-683-5769; e-mail: puborder@headstartinfo.org; Web site: http://www.headstartinfo.org. For full text: http://www.acf.hhs.gov/ programs/core/ongoing_research/faces/faces2000_final.pdf. |
| EDRS PRICE | EDRS Price MF01/PC10 Plus Postage. |
| DESCRIPTORS | Classroom Environment; Client Characteristics (Human Services); *Educational Quality; Family Characteristics; Family Influence; Kindergarten Children; Longitudinal Studies; *Outcomes of Education; Parent Background; Predictive Validity; *Preschool Children; Preschool Curriculum; *Preschool Education; *Program Effectiveness; Program Evaluation; Social Development; Student Characteristics; Teacher Characteristics; Test Validity |
| IDENTIFIERS | *Head Start Family and Child Experiences Survey; Head Start Program Performance Standards; Program Characteristics; *Project Head Start |

ABSTRACT

In 1997, Head Start launched the Family and Child Experiences Survey (FACES), a study of a national random sample of Head Start programs, designed to describe characteristics, experiences, and outcomes for children and families served by Head Start. In 2000, FACES began data collection on a new national cohort of 2,800 children and their families in 43 Head Start programs. This report focuses on the 2000 sample, as well as on Kindergarten followup data from the 1997 sample. For the 2000 sample, there were 4 phases of data collection following 3- and 4-year-olds from program entry, through 1 or 2 years of program experience to the Spring of their kindergarten year. Data collection included child assessment, parent interviews, teacher and staff interviews, and classroom observations. The report presents findings with regard to Head Start children's cognitive and social-emotional development, Head Start programs' use of curricula, quality in Head Start classrooms, relationship of program and classroom characteristics to children's outcomes, relationship of family and parental characteristics to children's outcomes, and predictive validity of the FACES cognitive and behavioral measures. Findings indicated that children entered Head Start at a

Reproductions supplied by EDRS are the best that can be made
from the original document.

great disadvantage to other children. Although the gap between Head Start children and the general population of preschool children narrows during the Head Start year on key components of school readiness, Head Start children enter kindergarten still substantially below national averages. Head Start classrooms continued to be of good quality across a wide variety of indicators. Several program factors were related to child outcomes, including higher teacher salaries and the use of an integrated curriculum. Head Start involvement moderated negative effects of violence, depression, and other risk factors on children's cognitive and social-emotional status. (Contains 21 references.) (KB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 478 791

# Head Start FACES 2000:
# A Whole-Child Perspective on Program Performance
## Fourth Progress Report

May 2003

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Child Outcomes Research and Evaluation
and the Head Start Bureau
Administration for Children and Families
U.S. Department of Health and Human Services

PS 031433

BEST COPY AVAILABLE

# Head Start FACES 2000:
# A Whole-Child Perspective on Program Performance

May 2003

| | |
|---|---|
| Prepared for: | Louisa B. Tarullo, Ed.D.<br>Child Outcomes Research and Evaluation<br>Office of Planning, Research and Evaluation<br>Administration for Children and Families<br>U.S. Department of Health and Human Services<br>Washington, DC |
| Prepared by: | |
| Westat: | Nicholas Zill, Ph.D., Project Director<br>Gary Resnick, Ph.D., Associate Project Director<br>Kwang Kim, Ed.D., Senior Research Analyst<br>Kevin O'Donnell, M.P.P., Research Analyst<br>Alberto Sorongon, Research Associate |
| Xtria: | Ruth Hubbell McKey, Ph.D., Co-Project Director<br>Shefali Pai-Samant, Ph.D., Senior Research Analyst<br>Cheryl Clark, M.Ed., Senior Research Analyst |
| The CDM Group, Inc: | Robert O'Brien, Ph.D., Co-Project Director<br>Mary Ann D'Elio, M.S., Senior Research Analyst |

This report was prepared for the Administration for Children and Families, U.S. Department of Health and Human Services (DHHS) under contract HHS-105-96-1912, Head Start Quality Research Consortium's Performance Measure Center.

# Acknowledgments

This report would not have been possible without the commitment and cooperation of the staff and families at the Head Start programs in both the 1997 and 2000 cohorts of FACES. Staff responded thoughtfully to interviews, welcomed us into their classrooms, and helped us to arrange interviews with families and assessments with children. We are also grateful to the Head Start families who patiently met with us and provided insights into their time with Head Start and beyond. The report also benefited from the advice and comments of the members of the Head Start Quality Research Centers Consortium, who served as technical advisers to the study and provided valuable feedback on earlier drafts of this report.

## Head Start FACES 2000:
## A Whole-Child Perspective on Program Performance

## Table of Contents

**Executive Summary**

**Introduction**
- A. The Head Start Program Performance Measures Initiative
- B. Design and Methodology
- C. Related Research Initiatives
- D. Overview of Report

**Chapter I: Cognitive Gains Made by Head Start Children and Their Achievement in Kindergarten**
- Research Questions
- Findings
  - A. Emergent Literacy and Mathematics Skills of Head Start Children Compared With Those of the General Population of Preschoolers
  - B. Change in Knowledge and Skills Over the Head Start Year
  - C. Emergent Literacy Gains of Language-Minority Children
  - D. Growth of Knowledge and Skills in Kindergarten
- Conclusions

**Chapter II: Social Skills and Problem Behavior of Head Start Children and Their Adjustment to Kindergarten**
- Research Questions
- Findings
  - A. Cooperative Classroom Behavior of Head Start
  - B. Problem Behavior of Head Start Children
  - C. Behavior and Adjustment in Kindergarten

**Chapter III: Relationship Between Curricula and Family, Program, and Classroom Characteristics**
- Research Questions
- Methods
- Findings
  - A. Head Start Teachers Report Using a Specific Curriculum, Receiving Training, and Having Access to Their Curriculum Methods
  - B. Relationship Between Curricula and Family, Program, and Classroom Characteristics
- Implications

**Chapter IV: Understanding Quality in Head Start Classrooms**
- Research Questions
  - A. What Quality Means and How We Appraise It
- Findings
  - B. The Quality of Most Head Start Classrooms Continues To Be Good
  - C. Teacher Qualifications and Experience Are Related to Classroom Quality
  - D. Teacher Backgrounds and Experience Are Related to Their Attitudes and Knowledge
  - E. Teacher Attitudes and Knowledge Are Related to Classroom Quality in Head Start
  - F. Teacher and Program-Level Factors Explain Significant Variations in Classroom Quality
  - G. Summary
- Key Findings

**Chapter V: Relationship of Program and Classroom Characteristics to Children's Cognitive Gains and Social Development in Head Start**
- Conceptual Framework
- Research Questions
- Analysis Method
- Findings
  - A. Use of an Integrated Curriculum
  - B. Higher Teacher Salary Levels
  - C. Teachers With Bachelors' or Associates' Degrees
  - D. Full-Day Versus Part-Day Classes
  - E. More Frequent Parental Reading to Children
  - F. Classroom Quality Indicators Not Found to Relate to Gains
- Summary and Discussion

**Chapter VI: Relationship of Child and Family Characteristics to Children's Cognitive Gains and Social Development in Head Start**
- Research Questions
- Findings
  - A. Prevalence of Family and Parental Risk Factors and Their Relationship to Children's Outcomes
  - B. Family Involvement and its Relationship to School Readiness
  - C. Head Start's Protective Role for Families and Children
- Summary

**Chapter VII: Predictive Validity of Cognitive and Behavioral Measures: Relationships Within and Across Cognitive and Social Developmental Domains**
  Research Questions
  Findings
    A. Children's Scores on the FACES Instruments at the End of Head Start Predict Kindergarten Outcomes
    B. FACES Behavioral Ratings at the End of Head Start Predict Children's Social Competence in Kindergarten
    C. Behavior Ratings at the End of Head Start Predict Reading Skills and General Knowledge at the End of Kindergarten
    D. Head Start Fall to Spring Gain Scores From the FACES Battery Predict Kindergarten Outcomes
    E. Head Start Fall to Spring Gain Scores From the Behavior Ratings Predict Social Competence in Kindergarten
    F. Head Start Fall to Spring Gain Scores From the Behavior Ratings Predict Reading Skills and General Knowledge at the End of Kindergarten
    G. Head Start Fall to Spring Gain Scores From the Behavior Ratings Contribute to the Prediction of Reading and General Knowledge at the End of Kindergarten
  Conclusions

**References**

**Appendix**

# List of Figures

**Introduction**

Figure I.1 Head Start Program Performance Measures Conceptual Framework

**Chapter I**

Figure 1.1 Most Children Entering Head Start Have Academic Skills Below National Norms
Figure 1.2 Head Start Students Show Gains in Vocabulary and Early Writing Skills During Program Year
Figure 1.3 Vocabulary Standard Scores of Children in Fall and Spring of Head Start Year: FACES 1997 Versus FACES 2000
Figure 1.4 Early Writing Standard Scores of Children in Fall and Spring of Head Start Year: FACES 1997 Versus FACES 2000
Figure 1.5 Letter Identification Standard Scores of Children in Fall and Spring of Head Start Year: FACES 1997 Versus FACES 2000
Figure 1.6: Early Math Standard Scores of Children in Fall and Spring of Head Start Year: FACES 1997 Versus FACES 2000
Figure 1.7 Children who Enter Head Start With Lower Skills Show Larger Gains in Comparison to National Norms
Figure 1.8 English-Language Vocabulary Skills of Head Start Children: Spanish-Speaking Language-Minority, Language-Majority, and Combined Population
Figure 1.9 Letter-Identification in English by Head Start Children: Spanish-Speaking Language Minority, Language-Majority, and Combined Population
Figure 1.10 Vocabulary and Letter Identification Skills in English and in Spanish of Head Start Children From Spanish-Speaking Language-Minority Families
Figure 1.11 Head Start Graduates Show Further Progress Toward National Norms in Kindergarten

**Chapter II**

Figure 2.1 Head Start Students Show Gains in Cooperative Classroom Behavior During Program Year
Figure 2.2 Head Start Students Show Significant Reductions in Hyperactive Behavior During Program Year
Figure 2.3 Head Start Students Showed Similar Patterns of Problem Behavior in 2000-2001 and 1997-1998 Program Years
Figure 2.4 Head Start Students who enter With High Levels of Problem Behavior Show Improvement During Program Year
Figure 2.5 Parents of Head Start Children Report Reductions in Problem Behavior During Program Year
Figure 2.6 Parents Report Greater Improvement in Head Start Students who Enter With High Levels of Problem Behavior

**Chapter III**

Figure 3.1 The Majority of Teachers Used Either Creative Curriculum or High/Scope
Figure 3.2 Teachers Liked Their Respective Curricula for a Variety of Reasons
Figure 3.3 Most Teachers From Northeast, Midwest, and West Used Creative Curriculum or High/Scope While Majority of Teachers From the South Used Other Curricula
Figure 3.4 The Classrooms of Teachers Who Used Creative Curriculum or High/Scope had Significantly Higher Average ECERS-R Total Scores and ECERS-R Language Scores Than Classrooms of Teachers Who Used Other Curricula
Figure 3.5 The Classrooms of Teachers who Used Creative Curriculum and High/Scope had Significantly Higher Average Quality Composite Scores Than Classrooms Using Other Curricula
Figure 3.6 Classrooms That Used Creative Curriculum and High/Scope had Significantly Higher Average ECERS-R Total Scores and ECERS-R Language Scores Than Classrooms That Used All Other Curricula. Classrooms That Used Widely Available Curricula Had Significantly Higher Average ECERS-R Language Scores Than Classrooms That Used All Other Curricula
Figure 3.7 Classrooms That Used Creative Curriculum and High/Scope had Significantly Higher Average Quality Factor Scores Than Classrooms That Used All Other Curricula. There Were no Significant Differences in the Average Quality Composite Scores Between Classrooms That Used Widely Available Curricula and All Other Curricula

**Chapter IV**

Figure 4.1 Classroom Quality in Head Start Compared With Other Preschool and Child Care Settings
Figure 4.2 Distribution of Classrooms on ECERS and ECERS-R, Fall 1997 and 2000
Figure 4.3 Teacher Beliefs Are Correlated With Classroom Quality, Fall 2000
Figure 4.4 A Model of Program- and Classroom-Level Factors Explaining Quality in Head Start

**Chapter V**

Figure 5.1 Analytical Model of Multi-Level Factors Predicting to Classroom Quality and Children's Achievement and Gains in the Head Start Year
Figure 5.2 Children in Head Start Program Using High/Scope Curriculum Show Greater Gains in Letter Recognition Skills
Figure 5.3 Children in Head Start Programs With Higher Teacher Salaries Show Larger Gains in Cooperative Classroom Behavior
Figure 5.4 Children in Head Start Programs With Higher Teacher Salaries Show Larger Declines in Hyperactive Behavior
Figure 5.5 Children in Head Start Classes Taught by Teachers With Bachelors' or Associates' Degrees Show Gains in Early Writing Skills
Figure 5.6 Head Start Children in Full-Day Classes Show Larger Gains in Letter Recognition Skills Than Those in Part-Day Classes

Figure 5.7 Children in Full-Day Head Start Classes Show Greater Gains in Early Writing Skills
Figure 5.8 Head Start Children Whose Parents Read to Them More Often Show Larger Vocabulary Gains
Figure 5.9 Children in Head Start Classes With Higher, Moderate, and Lower Quality Language Activities Show Parallel Gains in Vocabulary Knowledge
Figure 5.10 Children in Head Start Classes With Higher Child-Adult Ratios Show Larger Gains in Letter Recognition

**Chapter VII**

Figure 7.1 Correlations and Standardized Multiple Regression Coefficients Between Reading Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start Year
Figure 7.2 Correlations and Standardized Multiple Regression Coefficients Between General Knowledge Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start Year
Figure 7.3 Correlations and Standardized Multiple Regression Coefficients Between Reading Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start Year—Norm-Referenced Tests Only
Figure 7.4 Correlations and Standardized Multiple Regression Coefficients Between General Knowledge Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start Year—Norm-Referenced Tests Only
Figure 7.5 Correlations and Standardized Multiple Regression Coefficients Between Reading Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start Year—Criterion-Referenced Tests Only
Figure 7.6 Correlations and Standardized Multiple Regression Coefficients Between General Knowledge Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start Year—Criterion-Referenced Tests Only
Figure 7.7 Correlation Coefficients Between Parent and Teacher Ratings of Children's Social Competence at the End of Head Start and Teacher Reported Ratings of Cooperative Classroom Behavior at the End of Kindergarten
Figure 7.8 Correlations and Significant Standardized Multiple Regression Coefficients Between Teacher Ratings of Cooperative Classroom Behavior at End of Kindergarten Year and Parent and Teacher Behavior Ratings at End of Head Start Year
Figure 7.9 Correlations and Significant Standardized Multiple Regression Coefficients Between Teacher Ratings of Total Problem Behavior at End of Kindergarten Year and Parent and Teacher Behavior Ratings at End of Head Start Year
Figure 7.10 Correlation Coefficients Between Parent and Teacher Ratings of Children's Social Competence at the End of Head Start and Reading Scale Scores at End of Kindergarten
Figure 7.11 Correlations and Significant Standardized Multiple Regression Coefficients Between Reading Scale Scores at End of Kindergarten Year and Parent and Teacher Behavior Ratings at End of Head Start Year

Figure 7.12 Correlation Coefficients Between Parent and Teacher Ratings of Children's Social Competencies at the End of Head Start and General Knowledge Scale Scores at End of Kindergarten
Figure 7.13 Correlations and Significant Standardized Multiple Regression Coefficients Between General Knowledge Scale Scores at End of Kindergarten Year and Parent and Teacher Behavior Ratings at End of Head Start Year
Figure 7.14 Correlations and Standardized Multiple Regression Coefficients Between Reading Scale Scores at End of Kindergarten Year and FACES Assessment Scale Gain Scores During Head Start Year
Figure 7.15 Correlations and Standardized Multiple Regression Coefficients Between General Knowledge Scale Scores at End of Kindergarten Year and FACES Assessment Scale Gain Scores During Head Start Year
Figure 7.16 Correlations and Standardized Multiple Regression Coefficients Between Teacher Ratings of Cooperative Classroom Behavior at End of Kindergarten Year and Parent and Teacher Behavior Ratings Gain Scores During Head Start Year
Figure 7.17 Correlations and Standardized Multiple Regression Coefficients Between Teacher Ratings of Problem Behavior at End of Kindergarten Year and Parent and Teacher Behavior Ratings Gain Scores During Head Start Year
Figure 7.18 Correlations and Odds-Ratio Estimates Between Teachers' Decisions at End of Kindergarten Year to Assign Child to Repeat Kindergarten and FACES Assessment Scale Scores at End of Head Start Year
Figure 7.19 Correlations and Odds-Ratio Estimates Between Teachers' Decisions at End of Kindergarten Year to Assign Child to Repeat Kindergarten and Parent and Teacher Behavior Ratings at End of Head Start Year

## List of Tables

**Chapter III**

Table 3.1 The Majority of Urban Teachers Used Either Creative Curriculum or High/Scope While the Majority of Rural Teachers Used Creative Curriculum
Table 3.2: Type of Curricula and Classroom Quality Scores

**Chapter IV**

Table 4.1 Comparison of Lead Teacher Backgrounds, FACES Fall 1997 and Fall 2000
Table 4.2 Summary of Center- and Classroom-Level Factors Predicting Classroom Quality, Fall 2000

**Chapter V**

Table 5.1 Three-Level Regression Models of Assessment Scale Scores of Head Start Children in Fall and Spring of Program Year, and Fall-Spring Gains, 2000-2001
Table 5.2 Three-Level Regression Models of Assessment Scale Scores of Head Start Children in Fall and Spring of Program year, and Fall-Spring Gains, 2000-2001

**Chapter VI**

Table 6.1 Significant Head Start Moderators of Negative Outcomes Associated With Risk Factors
Table 6.2 Significant Adjusted Odds Ratios and 95% Confidence Intervals for Maternal Depression, Exposure to Violence, Domestic Violence, Alcohol Use, and Involvement With the Criminal Justice System

# Head Start FACES 2000:
# A Whole-Child Perspective on Program Performance

In 1997, Head Start launched the Family and Child Experiences Survey (FACES), a study of a national random sample of Head Start programs designed to answer critical questions about child outcomes and program quality. In 2000, FACES began data collection on a new national cohort—FACES 2000—and plans are underway for a third cohort. Now, longitudinal data on successive, scientifically representative samples of children, families, teachers, classrooms, and programs are available.

In both studies, children entered Head Start at a great disadvantage to other children, as evidenced by the children's initial scores on standardized assessments of cognitive skills. Findings from both cohorts of FACES show that the gap between Head Start children and the general population of preschool-age children narrows during the Head Start year on key components of school readiness. This is true to a greater extent in the 2000-2001 program year. However, despite the gains they make, Head Start children enter Kindergarten still substantially below national averages on such assessments.

Children made significant gains during the Head Start year relative to national norms, most notably in the areas of vocabulary knowledge and early writing skills. In the areas of letter recognition and knowledge of book and print conventions, children in 2000-2001 made significantly greater gains than Head Start children in 1997-1998. Gains in vocabulary and early writing were similar to those in 1997-1998. In both cohorts, children who entered Head Start with lower skill levels made greater gains than those who entered with higher skill levels. This finding may be related in part to the tendency of scores to move closer to the population mean over successive assessments.

In the domain of social and emotional development, children also showed growth in social skills and reduction in hyperactive behavior during the Head Start year. Children with high levels of shy, aggressive, or hyperactive behaviors (scoring in the top quarter) showed significant reduction of these behaviors.

FACES 2000 also found that Head Start classrooms continue to be of good quality across a wide variety of indicators. In 2000-2001 there was an increased percentage of new teachers with advanced degrees entering Head Start, compared to 1997-1998. Head Start teachers in FACES 2000 were more likely to be younger, new to teaching Head Start and entering with higher educational levels including graduate degrees. They were also more likely to be trained in Early Childhood Education and to be members of a professional organization. Results also show that Head Start teachers with higher levels of educational attainment, and with more years of teaching experience overall, were more likely to have knowledge and positive attitudes about early childhood education practices, which subsequently influenced classroom quality. Thus, teacher attitudes and knowledge mediates the relationship between teacher education and classroom quality.

The vast majority of programs use a specific curriculum, as mandated by the Head Start Program Performance Standards. Several program factors are related to child outcomes, including higher teacher salaries and use of an integrated curriculum. Teachers' educational credentials are linked to greater gains in early writing skills. In addition, provision of preschool services for a longer period each day is linked to greater cognitive gains.

FACES 2000 presents information on family and parent characteristics important to any investigation of school readiness. For example, when parents read more frequently to their children, their children had higher scores on early literacy assessments. In addition, when families engaged their children in weekly and monthly educational activities, there was a positive correlation with positive child behaviors and emergent literacy. Head Start families face numerous risks and challenges that are related to children's well-being. When Head Start parents were more depressed, their children had lower scores on a variety of cognitive measures. Positive correlations were also found between increased exposure to neighborhood violence and reports of child problem behavior. Parent involvement in Head Start was positively correlated with a number of positive cognitive and social outcomes. Importantly, Head Start involvement by parents moderated the negative effects of violence, depression, and other risk factors on children's cognitive and social-emotional status.

**The Study**

FACES describes the characteristics, experiences, and outcomes for children and families served by Head Start. It also explores the relationships among family and program characteristics and outcomes. In 1997, the FACES design included a nationally representative sample of 3,200 children and their families in 40 programs. The FACES 2000 sample includes 2,800 children and their families in 43 different Head Start programs across the nation. The current report focuses on the FACES 2000 sample, as well as Kindergarten follow-up data from the FACES 1997 sample.

Each cohort of FACES employs a nationally stratified sample of Head Start programs, centers, classrooms, children, and parents. FACES 2000 features four phases of data collection and follows 3- and 4-year-old children from program entry, through one or two years of program experience, with follow-up in the spring of Kindergarten. The FACES 2000 battery has four main components: the direct child assessment, parent interview, teacher and staff interviews, and classroom observations. Although there is no non-Head Start comparison group in FACES, the use of assessment measures with national norms permits comparisons between the skills of children in the sample and children of the same ages in the norming samples. Child outcomes can be compared with national averages on a range of standardized assessments with a mean of 100, and standard deviation of 15.

**Study Findings**

FACES provides information about the knowledge and skills that children have when they enter the Head Start program and the gains they make during the Head Start year and the first year of elementary school. It also describes the quality of Head Start classrooms, and factors that help explain variations in quality across Head Start classrooms. In addition, FACES 2000 data provide insight into the relationship of program and classroom characteristics to children's outcomes, as well as the relationship of family and parental characteristics to children's outcomes.

**Head Start Children's Cognitive and Social-Emotional Development**

- Most children entered Head Start at a great disadvantage, with early literacy and math skills substantially below national averages. The typical Head Start child was found to enter at about the 16th percentile in vocabulary and early writing skills, at about the 31st percentile in letter recognition and at about the 21st percentile in early math, when compared to the full spectrum of American children in the same age range.

- There was considerable diversity in skill levels among Head Start children, however. The highest quarter of Head Start children were at or above the national average (50th percentile) in early language and number skills, while the lowest quarter of children ranked on average in the lowest 2 percent of all U.S. preschoolers in these areas.

- As in 1997-1998, the gap between Head Start children and other preschool-age children narrowed during the Head Start year, especially with respect to vocabulary knowledge and early writing skills. Despite these gains, Head Start children still trail in these measures compared to national averages.

- Head Start children showed greater progress in letter recognition skills than they had in 1997-1998, but they still did not reach national averages in this area. Although the children made progress in early math skills, they did not make gains toward national averages in this domain.

- In 2000, Head Start children entered the program knowing about 4 letters of the alphabet, and left the program knowing about 9 letters on average, close to the congressional mandate of being able to name at least 10 letters.

- Children who entered the program with lower levels of knowledge and skill showed larger gains during the program year, yet still lagged considerably behind national averages. Children who started with higher assessment scores in the fall wound up with higher scores in the spring, but showed less dramatic gains. The finding of greater gains for children who entered with lower scores may be related in part to the tendency of scores to move closer to the population mean over successive assessments.

- Spanish-speaking children in Head Start showed significant gains in English vocabulary skills without declines in their Spanish vocabulary skills. They did not gain in letter recognition skills.

- Based on follow-up of the 1997-1998 cohort, Head Start graduates showed further progress toward national averages during kindergarten. Gains of between a third to more than half a standard deviation were observed in vocabulary, early math, and early writing skills during kindergarten. Most Head Start graduates could identify most or all of the letters of the alphabet by the end of kindergarten and more than half could recognize beginning sounds of words. Nevertheless, Head Start graduates remained behind their more advantaged peers in early achievement.

- The size of gains that children made while in Head Start were predictive of their achievement levels by the end of kindergarten.

- Children showed growth in social skills and reduction in hyperactive behavior during the Head Start year. According to teacher report, the average score of Head Start children on a cooperative classroom behavior rating scale increased significantly from fall to spring. In addition, the average score of Head Start children on a hyperactive behavior rating scale decreased significantly during the program year.

- Children with high levels (scoring in the top quarter) of shy, aggressive, or hyperactive behavior showed significant reductions in these problem behaviors in Head Start.

- Behavior in Head Start is predictive of adjustment and performance in early elementary school. Cooperative classroom behavior ratings and problem behavior ratings by Head Start teachers of children at the end of Head Start were predictive of behavioral adjustment ratings by kindergarten teachers in the spring of the kindergarten year. In addition, children who received higher cooperative behavior ratings and lower problem behavior ratings from Head Start teachers scored better on cognitive assessments at the end of Kindergarten, even when their test scores in Head Start were taken into account.

**Head Start Programs' Use of Curricula**

- The great majority of Head Start programs use a curriculum, as mandated by the Head Start Program Performance Standards. A wide variety of curricula are used, with a majority of programs selecting an integrated curriculum such as Creative Curriculum or High/Scope.

- There is a relationship between program characteristics (region, urban-rural, characteristics of children and families) and the type of curriculum used.

- There is a relationship between curricula and classroom quality, which may reflect the influences of other factors (such as the resources available to programs for purchasing and training in specific curricula), or may demonstrate the effect of certain curricula on quality.

**Quality in Head Start Classrooms**

- Head Start quality has been observed to be consistently good over time, using a variety of indicators including child-adult ratio, teacher-child interactions, and classroom activities and materials. Few classrooms scored below minimal quality. In fact, FACES shows that Head Start has a better, more limited range of quality than that seen in child care centers and preschools in several other national studies.

- More Head Start teachers in 2000-2001 had obtained a graduate school degree compared to 1997-1998. However, Head Start teachers have lower teaching qualifications on average than pre-kindergarten teachers in public elementary schools.

- Head Start teachers in 2000 are also younger, compared with those in 1997-1998, and more of them have been teaching in Head Start for two years or less. These newer teachers are also the ones most likely to have a graduate school degree.

- More teachers in 2000, studied Early Childhood Education or Child Development for their highest degree, and more teachers belong to a national professional association for early childhood educators, compared with those in 1997-1998.

- Teacher backgrounds, qualifications, and experience are related to their attitudes and knowledge of early childhood development practices. Teachers with higher scores for positive attitudes and knowledge about early childhood education practices were more likely to have higher levels of educational attainment, have some graduate school education or higher, have more total years teaching, and belong to an early childhood education association.

- Classrooms with higher levels of quality have teachers with higher levels of education, experience, and positive attitudes and knowledge about early childhood education practice.

- Teacher education and attitudes are linked to classroom quality. The relationship between teacher education and classroom quality is explained by teacher's attitudes and knowledge of early childhood education practice. Teachers who are more educated have more positive attitudes and knowledge, which translates into higher levels of classroom quality.

- Variations in the quality of Head Start classrooms may be explained by characteristics of the families and children they serve, by the curriculum used in the program, and by teacher attitudes and knowledge about early childhood education practice. Head Start programs that provide for a common integrative curriculum across classrooms and that pay their teachers better have sufficient resources available to positively influence classroom quality, through the quality of teachers hired, their experience and attitudes and knowledge.

## Relationship of Program and Classroom Characteristics to Children's Outcomes

- Higher teacher salaries are linked to greater gains in several cognitive and social-emotional areas, including letter identification, oral communication of basic social information, and cooperative classroom behavior. Children in programs with higher teacher salaries also showed greater improvement in hyperactive problem behavior during the Head Start year.

- Use of an integrated curriculum is linked to greater gains in several cognitive and social-emotional areas. Specifically, children in Head Start programs using High/Scope showed larger fall-spring gains in letter identification and cooperative classroom behaviors than children in programs using other curricula. Children in programs using High/Scope also showed greater improvement in total behavior problems and hyperactive problem behavior.

- Teachers' educational credentials are linked to greater gains in early writing skills. Children taught by Head Start teachers with Bachelors' Degrees or Associates' Degrees showed gains toward national averages in an assessment of early writing skills, whereas children taught by teachers with lesser credentials merely held their own against national norms.

- Provision of preschool services for a longer period each day is linked to greater cognitive gains. Children in full-day classes in Head Start showed larger fall-spring gains in letter recognition and early writing skills than did children in part-day classes.

- There is indirect evidence that encouraging parents to engage in more educational activities with their children at home is linked to greater cognitive gains. Children whose parents report reading to them every day show larger fall-spring gains in vocabulary knowledge and letter recognition skills than children whose parents report reading once or twice or less frequently per week.

- Within the generally good quality range of Head Start classrooms, variation in quality as measured by the ECERS-R Language scale or the Caregiver Interaction Scale is not associated with differences in fall-spring achievement gains across classes.

- Within the narrow range of group size in Head Start, variation in child/adult ratios is not associated with or is negatively associated with differences in fall-spring achievement gains across classes.

**Relationship of Family and Parental Characteristics to Children's Outcomes**

- Almost 90 percent of Head Start families manifested at least one of a set of six selected socioeconomic risk factors. About one fifth of the families had four or more risk factors. Children in these families had lower parent ratings on emergent literacy and higher teacher and parent ratings of problem behavior. In the assessments, these children scored lower on design copying, color naming, one-to-one counting, book knowledge, vocabulary, early math, early writing, letter identification, social awareness, comprehension, and print concepts.

- Twenty-five percent of the parents were classified as moderately or severely depressed. Parents who were more depressed reported that their children had more problem behaviors and fewer positive social behaviors. Their children also had lower scores on one-to-one counting, creativity, and early math assessments, after controlling for parent education, income, and other demographic factors.

- More than one fifth of the parents had witnessed violent crime. Five percent were victims of violent crime in the neighborhood, while a similar percentage were victims of violence in their homes. Parents reported that almost 10 percent of the children witnessed domestic violence during the previous year. Parents reported that less than 2 percent of the children had been victims of violent crime or victims of domestic violence. Positive correlations were found between increased exposure to violence and reports of child problem behavior and maternal depression, after controlling for parent education, income, and other demographic factors.

- Almost one fifth of the parents reported that someone in their household had been arrested and charged with a crime. Children in these families were more than three times more likely to have been a witness to either a violent crime or domestic violence in the past year. These children were also more than three times more likely to have been a victim of domestic violence or violent crime. These children had lower vocabulary scores, and were reported by both parents and teachers to be more aggressive and have more overall problem behaviors.

- Families engaged their children in a number of weekly and monthly educational activities. The number of activities was positively correlated with positive child behaviors and emergent literacy and negatively correlated with problem behaviors, after controlling for the number of times a child was read to in the past week, parent education, income, and other demographic factors. In particular, the weekly activities had positive correlations with scores on color naming and vocabulary. Monthly activities were positively correlated with the social awareness, color naming, one-to-one counting, book knowledge, and print concepts assessments.

- More than two thirds of parents had attended parent-teacher conferences, observed in their children's classrooms for at least 30 minutes, or met with a Head Start staff member in their homes. Parental involvement in Head Start was positively correlated with parental reports of positive social behavior and higher emergent literacy skills and negatively correlated with aggressive and overall problem behavior, after controlling for parent education, employment, and other demographic factors. Children with more involved parents scored higher on vocabulary, book knowledge, early writing, early math, and letter identification tasks.

- Preliminary findings suggest that Head Start may play a role in protecting children from the negative outcomes associated with family risk factors, including maternal depression, exposure to violence, alcohol use, and involvement in the criminal justice system. Parent involvement in Head Start, parent reports that they and their children had positive experiences at Head Start, and parent satisfaction with the program significantly moderated negative relationships between risk factors and child behavioral and cognitive outcomes, controlling for parent education, income, and employment; child age, gender, and race; and the overall family activity level with the child.

**Predictive Validity of the FACES Cognitive and Behavioral Measures**

- Children's scores on FACES assessments at the end of Head Start, as well as the gains they make during the Head Start year, strongly predict their performance at the end of Kindergarten. As an indicator of pre-literacy skills, the cognitive measures show strong associations with reading ability at the end of the kindergarten year. As an indicator of school adjustment and social competence, the behavior ratings demonstrate ability to predict kindergarten behaviors that promote learning and those that impede learning.

- The vocabulary and early literacy instruments used in FACES tap different types of abilities ("inside-out" or decoding skills as well as "outside-in" or comprehension skills) that are important for children's future reading proficiency and academic achievement.

- Scores from the FACES instruments and the FACES behavior ratings both contribute to the prediction of teachers' practical decision of whether a child repeats kindergarten or is promoted to first grade.

- The multi-measure and multi-method approach to the measurement of children's abilities provides a variety of information sources that significantly contribute to the prediction of Kindergarten outcomes.

# Head Start FACES 2000:
# A Whole-Child Perspective on Program Performance

## Introduction

As part of the Head Start Program Performance Measures Initiative, Head Start launched the Family and Child Experiences Survey (FACES) in Fall 1997. With a nationally representative sample of 3,200 children and their families in 40 programs, FACES describes the characteristics, experiences, and outcomes for children and families served by Head Start, as well as exploring the relationships among family and program characteristics and outcomes. In Fall 2000, Head Start began data collection on a new national cohort of FACES, called FACES 2000. The FACES 2000 sample includes 2,800 children and their families in 43 different Head Start programs across the nation.

FACES provides critical information for the Head Start program on important aspects of outcomes, quality and practices beyond the aggregated, administrative data previously collected. Through the ongoing, longitudinal FACES study, Head Start can examine key facets of program quality and children's school readiness on successive, scientifically representative samples of children, families, teachers, classrooms, and programs. While these data are crucial for decision-making at the national level, there are important limitations on the questions they can answer. They do not provide information on every child in each program, nor do they provide information on or comparisons to children recruited but not served by Head Start. Those questions are being answered via the Head Start National Reporting System and the Head Start Impact Study, respectively.

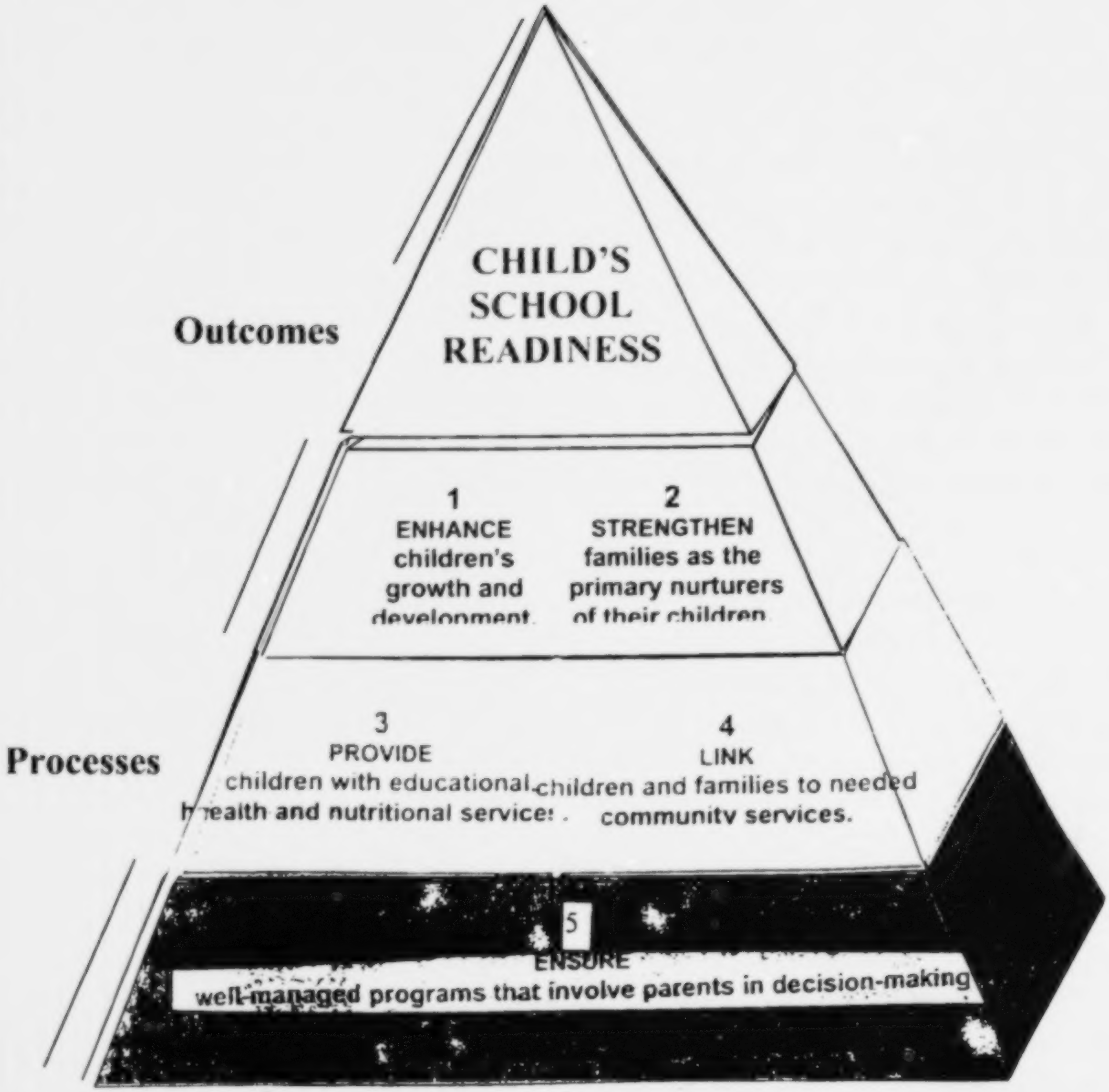
This introductory chapter describes the history and conceptual framework of FACES and the Head Start Program Performance Measures Initiative. In addition, it also discusses the research design and methodology of FACES, as well as the central information FACES provides in the context of related research endeavors. The chapter closes with an overview of the report.

### A. The Head Start Program Performance Measures Initiative

The Head Start Program Performance Measures, launched in 1995, were developed in accordance with the recommendations of the 1993 Advisory Committee on Head Start Quality and Expansion, the mandate of Section 641A (b) of the Head Start Act (42 USC 9831 et seq.) as reauthorized in 1994 and the Government Performance and Results Act (GPRA)(Public Law 103-62).

The Head Start Act defines Program Performance Measures as "methods and procedures for measuring, annually and over longer periods, the quality and effectiveness of programs operated by Head Start agencies" that will be used to identify strengths and weaknesses in the Head Start program--both nationally and by region--and identify programmatic areas requiring additional training and technical assistance.

Figure I.1 Head Start Program Performance Measures Conceptual Framework


Outcomes
CHILD'S SCHOOL READINESS
1
ENHANCE children's growth and development
2
STRENGTHEN families as the primary nurturers of their children
Processes
3
PROVIDE children with educational, health and nutritional services.
4
LINK children and families to needed community services.
5
ENSURE well-managed programs that involve parents in decision-making

BEST COPY AVAILABLE

## Conceptual Framework

In 1996-97, a conceptual framework for the Program Performance Measures was developed and the measures were finalized.[1] The conceptual framework unifies and organizes the Program Performance Measures to display the linkages between process and outcome measures for Head Start children and families. (See Figure 1.1 for the graphical representation of the framework.) The framework is based on the ultimate goal of Head Start, which is to promote the school readiness of children.

Head Start has adopted the "whole child" view of school readiness that was recommended by the Goal One Technical Planning Group of the National Education Goals Panel (Goal One Technical Planning Group, 1991, 1993). The panel defined school readiness as a multi-faceted phenomenon comprising five developmental domains that are important to the child's readiness for school: physical well-being and motor development, social and emotional development, approaches to learning, language usage and emerging literacy, and cognition and general knowledge. Each of these domains is represented in the battery of measures FACES uses to assess how well the national sample of Head Start programs is performing. It takes into account the interrelatedness of cognitive, emotional, and social development; physical and mental health; and nutritional needs. School readiness is depicted at the top of the pyramid, with five objectives supporting it:

- Objective 1. Enhance children's healthy growth and development.
- Objective 2. Strengthen families as the primary nurturers of their children.
- Objective 3. Provide children with educational, health and nutritional services.
- Objective 4. Link children and families to needed community services.
- Objective 5. Ensure well-managed programs that involve parents in decision-making.

Each of these objectives is critical to helping children of low-income families attain their full potential. They also represent the cornerstones of the Head Start program. Objectives 1 and 2 represent outcomes or results that the program is designed to produce. Achieving both of these objectives is critical to the ultimate success of Head Start. As parent involvement and family support are key tenets of Head Start, both child and family-oriented outcome measures are included here. Objectives 3, 4, and 5 comprise the lower tiers of the pyramid and contain the process measures that are the basis for the attainment of Objectives 1 and 2 and the ultimate goal of enhancing children's school readiness. An important aspect of the pyramid is the strong empirical connection between the provision of quality services (process measures) and improvements in child development (outcome measures).

---

[1] See *Head Start FACES: Longitudinal findings on program performance. Third progress report* (ACYF, 2001) for a list of the 24 Program Performance Measures.

## B. Design and Methodology

Each cohort of FACES employs a nationally representative sample of Head Start programs, centers, classrooms, children and parents.[2] The sample is stratified by three variables: region of the country (northeast, midwest, south or west); urbanicity (urban versus rural); and percentage of minority families in the program (50 percent or more vs. less than 50 percent). The first cohort of FACES had six phases of data collection, and followed 3- and 4-year-old Head Start children from entry into Head Start, through one or two years of program participation, with follow-up in spring of Kindergarten and spring of first grade (ACYF, 2001, 1998)[3]. FACES 2000 features four phases of data collection and follows 3- and 4-year-old Head Start children from program entry through spring of Kindergarten. For the current report, complete Kindergarten data are not yet available. However, the report includes information on Kindergarten performance from the FACES 1997 cohort.

### Survey Measures and Instruments

The FACES 2000 battery has four main components: the child assessment, parent interview, teacher and staff interviews, and classroom observations. The child outcomes include the major components of school readiness, and are collected through direct child assessments and rating scales completed by parents and teachers. Parent interviews are conducted with the primary caregiver of the Head Start child, and tap parenting behaviors, the socioeconomic characteristics of the family, and parental health and well-being. Interviews are administered to classroom teachers, center directors, program directors, and component coordinators to collect data on staff experience, education and training as well as knowledge and beliefs about child development, and educational activities with children and parents. Classroom observations collect data on both the structure of the classroom and classroom processes, such as teacher-child interactions. This battery has remained largely the same since 1997, with some minor revisions based on field experiences and newly released versions of instruments.[4]

## C. Related Research Initiatives

ACF considers FACES--and the ongoing information it provides about program quality and its relationship to child outcomes--a critical component of its effort to study Head Start. Indeed, policymakers relied on data from FACES during Head Start's 1998 Reauthorization, as Head Start officials were able to report to Congressional leaders on the quality of Head Start programs and the knowledge and skills of Head Start children as they completed the program. The data and experiences from FACES assisted the 1999 Advisory Committee on Head Start Research and Evaluation as it deliberated the design of a national impact study of Head Start as mandated by Congress. In addition, data from FACES have been widely disseminated within the Head Start community to assist with continuous efforts of program improvement and have guided training and technical assistance efforts.

---

[2] See the Appendix for details on sample design and response rates.

[3] The initial cohort of FACES also included an embedded case study of a longitudinal sample of 120 randomly selected families from the larger FACES sample; see *A Descriptive Study of Head Start Families: FACES Technical Report I,* (ACYF, 2002).

[4] See the Appendix for details on data collection instruments.

The purpose and design of FACES complement related ongoing research initiatives funded by ACF and other agencies interested in the well-being of children.

The National Head Start Impact Study is a longitudinal study that involves a sample of approximately 5,000 3- and 4-year old children across an estimated 75 nationally representative grantee/delegate agency groups. The Impact Study was congressionally mandated in the Head Start reauthorization of 1998. The Impact Study takes place only in communities where there are more eligible children and families than can be served by the program. This study has two primary goals. The first is to determine on a national basis how Head Start affects the school readiness of children participating in the program as compared to eligible children not enrolled in Head Start. The second goal of the study is to determine under which conditions Head Start works best and for which children. The children participating have been randomly assigned to either a treatment group (which receives Head Start services) or a comparison group (which does not receive Head Start services).

The Head Start Bureau is also conducting a field test of the National Reporting System on Child Outcomes (NRS). When fully implemented, the child outcomes information will assess the progress of approximately 500,000 4- and 5-year-old children in Head Start. It will produce a national outcomes report of children's ability and progress on the Presidentially and congressionally-mandated indicators.

The Head Start Quality Research Center (QRC) Consortium examines the efficacy of interventions designed to enhance the school readiness of preschool children in Head Start. These five-year grants fund partnerships between eight academic researchers and Head Start programs designed to improve child outcomes in the areas of literacy, social-emotional development, and other domains of school readiness, through enhancements to curriculum, teacher training and mentoring, parent involvement and assessment practices. Research teams are implementing and evaluating their interventions with Head Start program partners in an initial site and then replicating the successful interventions in additional sites. Cross-site data collection uses the FACES measures.

The Early Head Start Tracking/Pre-Kindergarten Follow-up of the Early Head Start Research and Evaluation project includes 17 local universities funded during the Birth to Three Phase. These sites will conduct cross-site and site-specific research, building upon earlier research and following the original children and families from the time they leave the Early Head Start program until they are ready to enter kindergarten.

ACF has also partnered with other federal agencies to study children in Head Start and other low-income children. One such effort is the ECLS Kindergarten Cohort, a longitudinal study of approximately 23,000 children nationwide who began kindergarten in the fall of 1998 and will be assessed through the fifth grade. An estimated 3,000 have been verified as former

Head Start children. In addition to contributing to this study, FACES utilizes measures developed in ECLS-K in its kindergarten follow-up.

In addition, ACF is partnering with the National Institute of Child Health and Human Development (NICHD), and the Assistant Secretary for Planning and Evaluation (ASPE) within the Department of Health and Human Services (DHHS), and the Office of Special Education and Rehabilitation Services (OSERS) of the U.S. Department of Education, to solicit grant applications to study the Effectiveness of Early Childhood Programs, Curricula, and Interventions in Promoting School Readiness. The purpose of these grants will be to study the effectiveness of integrative early childhood interventions and programs across a variety of early childhood settings in promoting school readiness for children, from birth through age five, who are at risk of later school difficulties.

## D. Overview of Report

This current document is the fourth progress report of FACES findings, and the first detailing findings from FACES 2000, the second full cohort of FACES. Subsequent chapters describe:

- The cognitive gains made by Head Start children in preschool and their achievement in kindergarten;
- Improvements in social skills and the amelioration of problem behavior in Head Start, and children's classroom adjustment and performance in kindergarten;
- The relationship between curricula and program, classrooms and child characteristics, including the types of curricula used by Head Start programs, sources of training, teacher access to and satisfaction with curricula, and the relationship between curricula and child, family, and program characteristics;
- The quality of Head Start centers and classes as early childhood care environments based on classroom observations, as well as the link between teacher education and beliefs and classroom quality, and among program resources and family characteristics and program quality;
- The relationship of program and classroom characteristics (including teacher salaries, use of an integrated curriculum, and traditional measures of classroom quality) to children's cognitive gains and social development in Head Start;
- The relationship between parent and family characteristics (including risk factors such as maternal depression, exposure to violence, domestic violence, substance use, and involvement in the criminal justice system) and children's social and cognitive outcomes, as well as the moderating effects of overall Head Start experiences.
- Predictive validity of cognitive and behavioral measures, and specifically relationships within and across cognitive and social developmental domains.

# REFERENCES

Administration on Children, Youth and Families. (2002). *A descriptive study of Head Start families: FACES technical report I.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth and Families. (2001). *Head Start FACES: Longitudinal findings on program performance. Third progress report.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth and Families. (1998). *Head Start Program Performance Measures: Second progress report.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth and Families. (1997). *First progress report on the Head Start Program Performance Measures.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth and Families. (1995). *Charting our progress: Development of the Head Start Program Performance Measures.* Washington, DC: U.S. Department of Health and Human Services.

Goal One Technical Planning Group. (1991). The Goal One Technical Planning Subgroup report on school readiness. In National Education Goals Panel (Ed.), *Potential strategies for long-term indicator development: Reports of the technical planning subgroups* (Report No. 91-0, pp. 1-18). Washington, DC: National Education Goals Panel.

Goal One Technical Planning Group. (1993). *Reconsidering children's early development and learning: Toward shared beliefs and vocabulary. Draft report to the National Education Goals Panel.* Washington, DC: National Education Goals Panel.

# Chapter I: Cognitive Gains Made by Head Start Children and Their Achievement in Kindergarten

The Head Start Family and Child Experiences Survey (FACES) provides information about the knowledge and skills that children who attend Head Start have when they enter the program and the gains they make during the Head Start year and the first year of elementary school. The information is helpful in assessing how well the Head Start program is performing, and what changes and reforms may be needed to improve program performance. The information is gained through direct, one-on-one assessment of nationally representative samples of Head Start students in the fall and spring of the program and at the end of their kindergarten year. Although there is no non-Head Start comparison group in FACES, the use of assessment measures with national norms permits comparisons between the skills of children in the sample and children of the same ages in the norming samples.[5]

## RESEARCH QUESTIONS

Assessment data from FACES 2000 were used to address the following research questions:

1. What skills and knowledge do children have when they enter Head Start programs?
2. Do children make significant gains in knowledge and skills during the Head Start year? During the kindergarten year?
3. How do these gains vary across skill areas and among children who enter the program with lower or higher knowledge levels?
4. Are the gains that Head Start children make changing? Did they change significantly between the 1997-1998 program year and the 2000-2001 year?

In making these comparisons, the analyses focused on children who were assessed in English in both the fall and spring of the Head Start year. Information is presented in a later section about the skills and knowledge of children who were initially assessed in Spanish because they came from Spanish-speaking homes and their knowledge of English was insufficient for testing in English in the fall.

## FINDINGS

Children who entered Head Start in the fall of 2000 had academic skill and knowledge levels well below national averages. They were comparable to the levels found in the initial round of FACES, conducted three years earlier in fall 1997. As in the earlier study, children made significant gains during the Head Start year, most notably in the areas of vocabulary knowledge and pre-writing skills. In the areas of letter recognition and knowledge of book and print conventions, children in 2000-2001 made significantly greater gains than Head Start children had in 1997-1998. In the areas of vocabulary and early math, gains were similar across the two studies. Children who entered Head Start with lower skill levels made greater gains than those who entered with higher skill levels, as was the case in the first round of FACES.

---

[5] This chapter focuses on cognitive measures with national norms. See Chapter II for information about social-emotional measures and the Appendix for a complete listing of measures used in FACES.

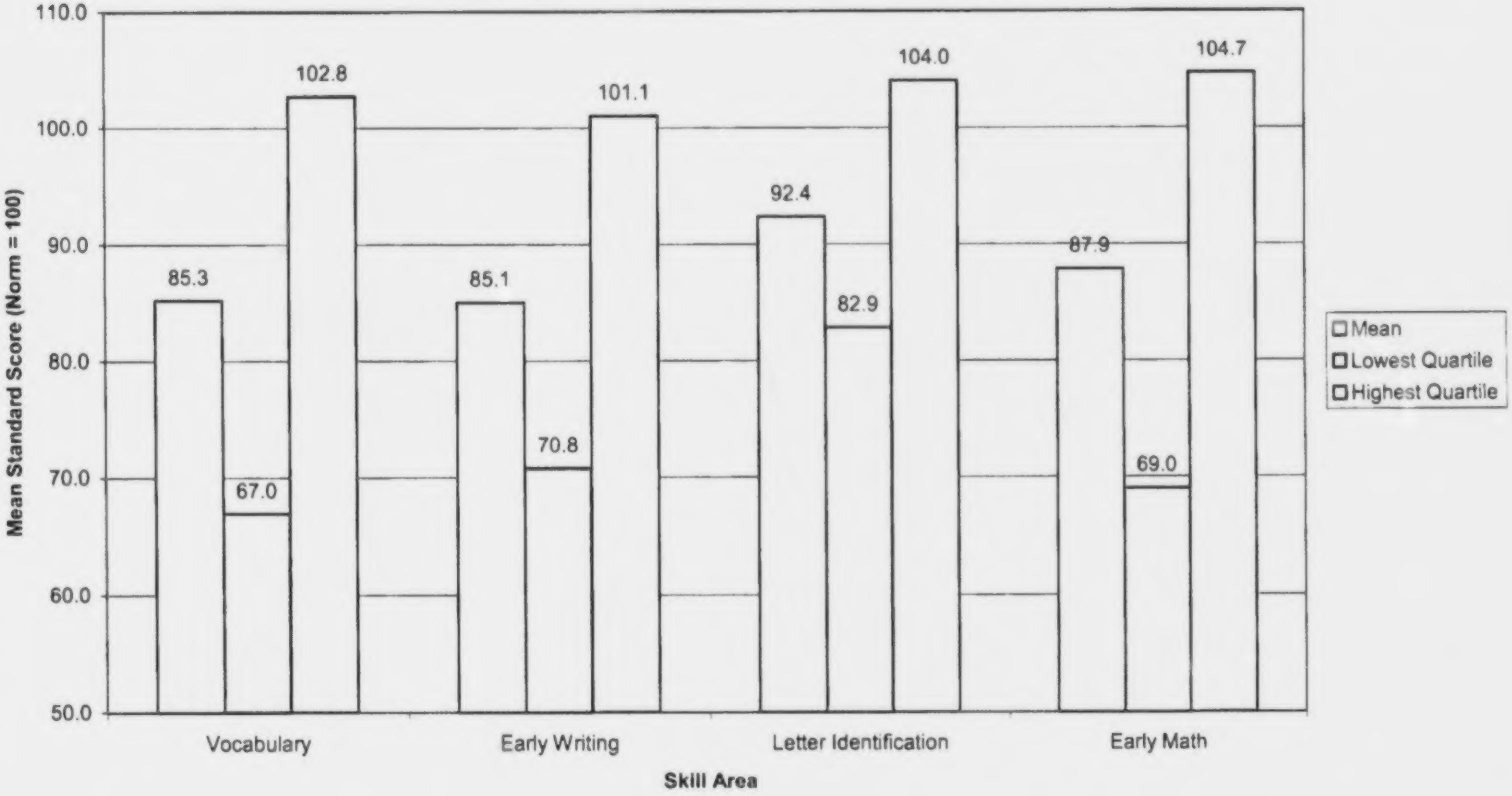
## A. Emergent Literacy and Mathematics Skills of Head Start Children Compared With Those of the General Population of Preschoolers

A primary focus of FACES was to measure the knowledge and skills children brought with them when they entered the Head Start program, and how this varied across academic skill areas. Cognitive measures with national norms available for comparison with the Head Start results included tests in vocabulary, early writing, letter identification, and early math.

### Majority Entered Head Start With Academic Skills Below National Norms

As in FACES 1997, the majority of children who entered Head Start in fall 2000 came into the program with early literacy and numeracy skills that were less developed than those of most children of the same age. This was to be expected with a group of young children who came from families with low parent education and income levels. The association between family socioeconomic status and children's achievement has often been demonstrated in education research (e.g., Phillips, Brooks-Gunn, Duncan, Klebanov, & Crane, 1998). FACES found that Head Start entrants had a mean standard score of 85.3 on the Peabody Picture Vocabulary Test (PPVT-III). They had mean standard scores of 85.1 on the Dictation (early writing) task of the Woodcock-Johnson Revised (WJ-R) achievement battery; 87.9 on the Applied Problems (early math) task, and 92.4 on the Letter-Word Identification (pre-reading) task. Standard scores are constructed to have an overall mean of 100 and a standard deviation of 15, and are based on a sample of children of a given age, across all income levels. Thus, the literacy and number skills that the average Head Start child brought to the program were from half a standard deviation to a full standard deviation below national averages. These scores imply that the typical Head Start child was at about the 16th percentile in vocabulary and early writing skills, at about the 31st percentile in letter recognition, and at about the 21st percentile in early math, when compared to the full spectrum of American children in the same age range.

**Figure 1.1. Most Children Entering Head Start Have Academic Skills Below National Norms**



SOURCE: Head Start Family and Child Experiences Survey (FACES), fall 2000, children assessed in English in both fall and spring.

### Diversity in Skills at Program Entry

Though most children had below-average literacy skills, FACES 2000 found considerable diversity in the Head Start population (Figure 1.1). For example, mean standard scores for the highest quarter of children entering Head Start were at national averages: 102.8 in vocabulary, 104.0 in letter recognition, 104.7 in early math, and 101.1 in early writing skills. Thus, these students would rank above the 50th percentile when compared to all U.S. preschoolers. On the other hand, mean standard scores for the lowest quarter of Head Start children were two standard deviations or more below national averages: e.g., 67.0 in vocabulary and 70.8 in early writing skills. These scores would rank the bottom quarter of Head Start students in the lowest 2 percent of all U.S. preschoolers. Similar diversity of skills was found in FACES 1997.

## B. Change in Knowledge and Skills Over the Head Start Year

A primary focus of FACES was to measure the extent of change in children's knowledge and skills from the fall to the spring of the Head Start year, especially in comparison to national averages for all children of the same ages. These changes were meant to serve as key indicators of the extent to which programs were enhancing children's school readiness.

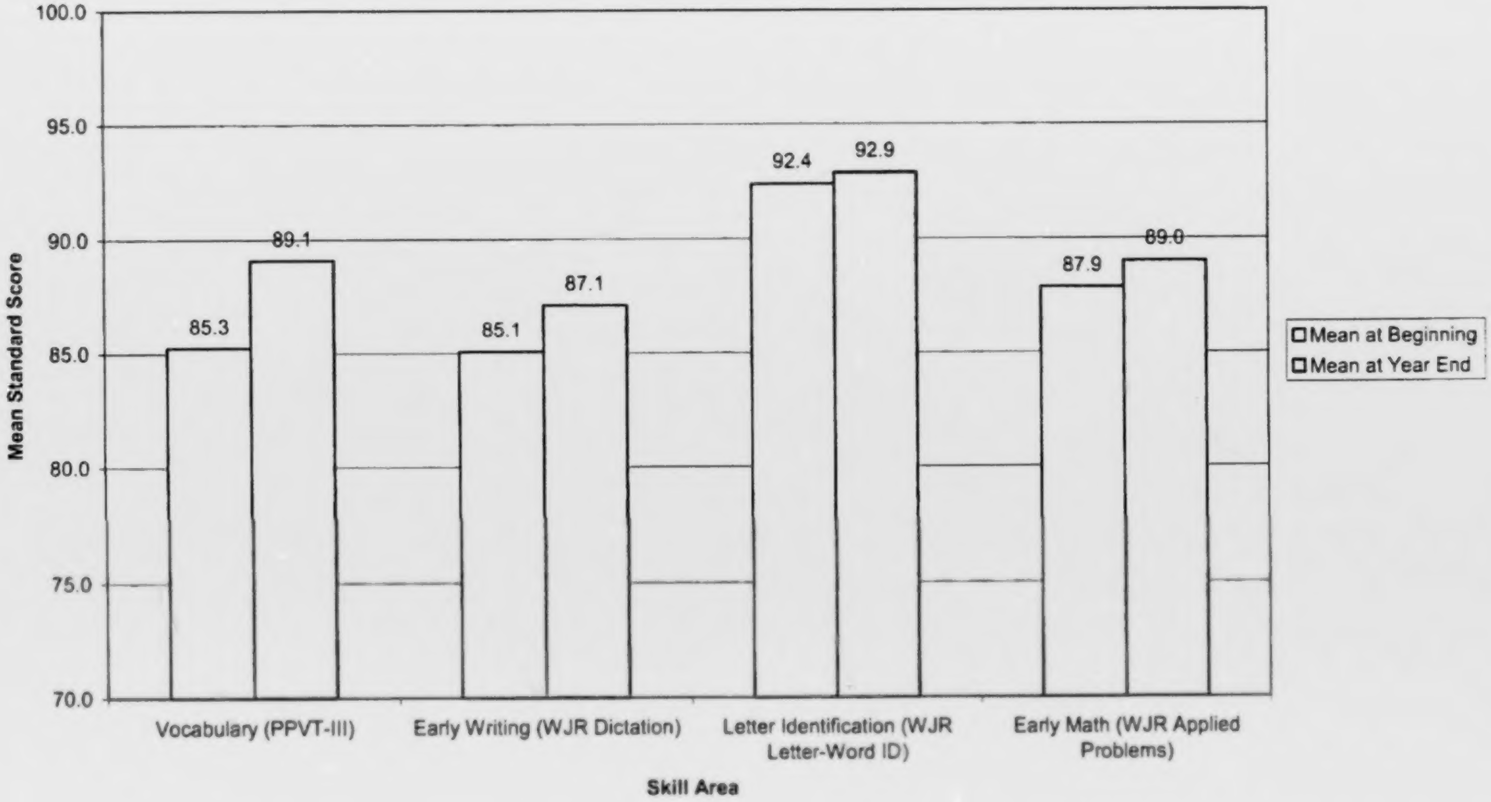
### Gains in Vocabulary Knowledge and Writing Skills

Children in Head Start showed significant expansion of their vocabularies between the beginning and end of the program year. By the spring of the Head Start year, mean standard scores were 89.1 for the PPVT-III and 87.1 on the WJ-R Dictation writing task. The mean standard score on the vocabulary test went up by 3.8 points ( $p < .001$), or more than one quarter of a standard deviation. The mean standard score on the writing task increased by 2.0 points ($p < .05$), or .15 of a standard deviation (Figure 1.2). The vocabulary gain seen in FACES 2000 was very comparable to that observed in FACES 1997, while the early writing gain, though still significant, was smaller ($p = .0028$)[6] (Figures 1.3, 1.4).

---

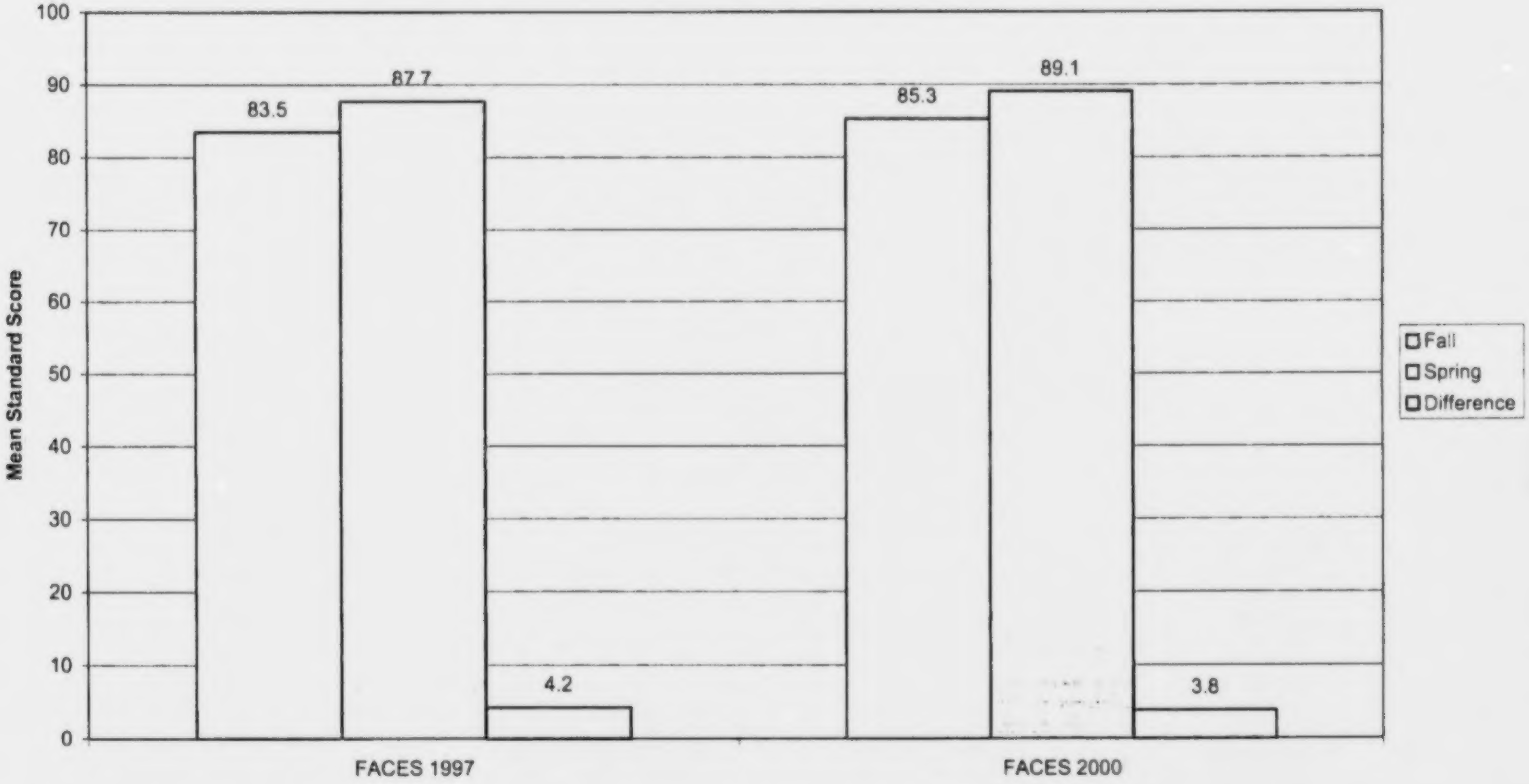
[6] The fall and spring mean scores and gains are those for children who were assessed in English in both fall and spring. These figures are comparable to those reported for FACES 1997. For information about the mean scores and gains of Spanish-speaking children from language-minority families, see below.

**Figure 1.2. Head Start Students Show Gains in Vocabulary and Early Writing Skills During Program Year**



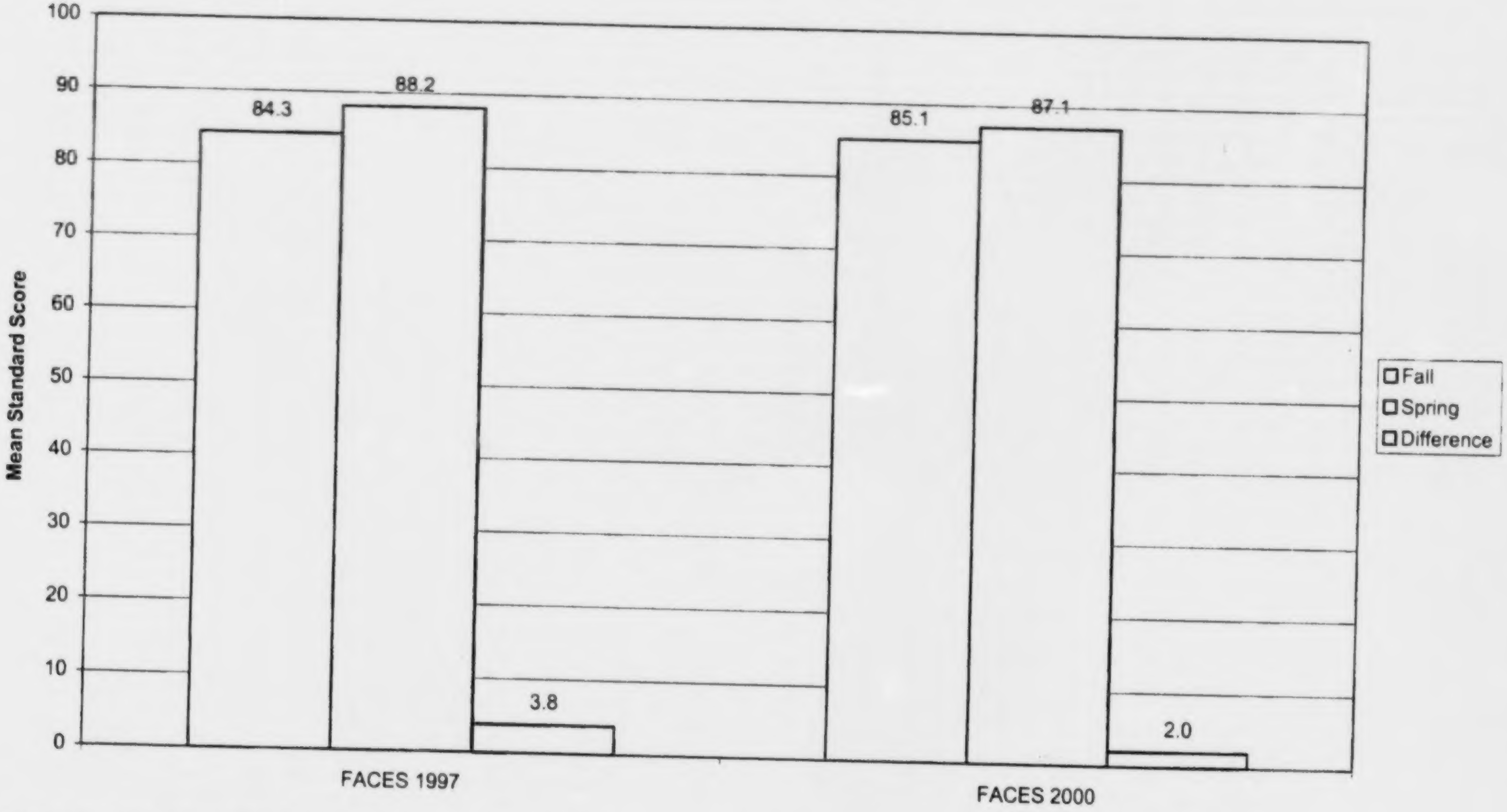
SOURCE: Head Start Family and Child Experiences Survey (FACES), fall 2000 and spring 2001, children assessed in English in both fall and spring.

**Figure 1.3. Vocabulary Standard Scores of Children in Fall and Spring of Head Start Year: FACES 1997 Versus FACES 2000**



SOURCE: Head Start Family and Child Experiences Survey (FACES), Fall 1997-Spring 1998, Fall 2000-Spring 2001, children receiving English language assessments both times.

**Figure 1.4. Early Writing Standard Scores of Children in Fall and Spring of Head Start Year: FACES 1997 Versus FACES 2000**



SOURCE: Head Start Family and Child Experiences Survey (FACES), Fall 1997-Spring 1998, Fall 2000-Spring 2001, children receiving English language assessment both times.

While the gains shown by Head Start children from fall to spring were relatively modest, they fell within the range that has been deemed "educationally meaningful" (Rosenthal & Rosnow, 1984). They were in line with earlier findings on the immediate effects of Head Start on children's intellectual performance (Haskins, 1989, p. 277; McKey et al., 1985). On the other hand, the vocabulary gains found in Head Start were about half the size of standard-score gains in IQ and achievement that have been obtained in some earlier studies of more intensive interventions with children from disadvantaged families (Barnett, 1998, pp. 13-14). Most children left Head Start with vocabulary knowledge and early writing skills that were still below national averages.

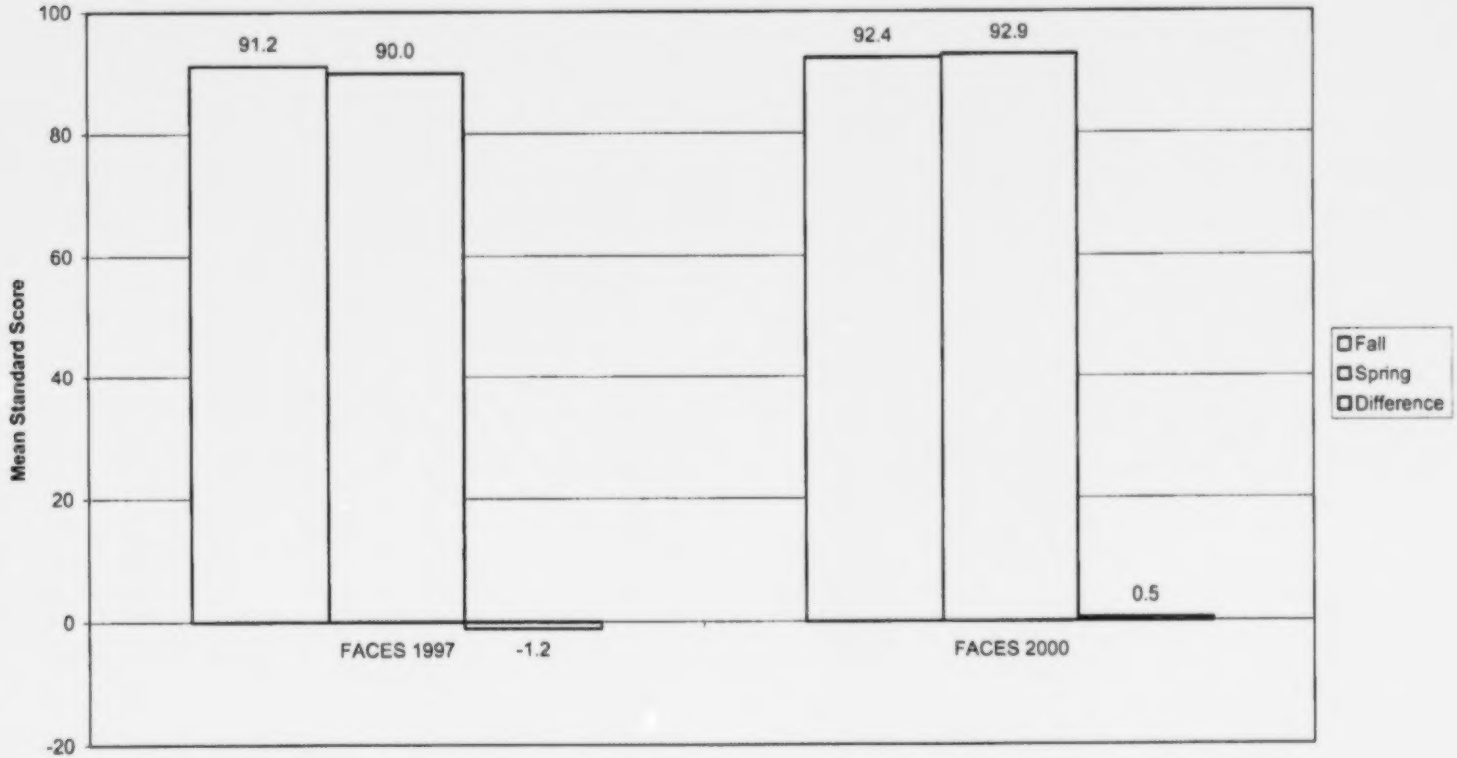
### Greater Gains in Letter Recognition and Book Knowledge

In recent years, the national Head Start program has been stressing the importance of early literacy skills like letter recognition. FACES 2000 found that Head Start children were making greater gains in letter recognition from fall to spring than was the case in 1997-1998. In fall 2000, the typical child was found to enter the program knowing about four letters (3.9) of the alphabet, and to leave knowing about nine (8.9). The equivalent gain observed in FACES 1997 was from about 3 letters (3.2) to seven letters (7.2).[7] This meant that children in the FACES 2000 cohort held their own in terms of standard scoregains on the WJ-R Letter-Word Identification task, whereas in FACES 1997, they actually showed a small decline in standard scores from fall to spring (Figure 1.5). The difference between the change observed in FACES 2000 and that observed in FACES 1997 was statistically significant ($p = .0005$).

---

[7] The mean number of letters known out of the entire English alphabet are estimates based on the mean scores children received on the WJ-R Letter-Word Identification (LWI) task. Estimates were derived using a known relationship between scores on the WJ-R LWI and a Letter Naming task in which children are shown all the letters of the alphabet. There are no national norms on the number of letters of the alphabet the typical four-year-old can name. For further details, see the Appendix .

Figure 1.5. Letter Identification Standard Scores of Children in Fall and Spring of Head Start Year: FACES 1997 Versus FACES 2000



SOURCE: Head Start Family and Child Experiences Survey (FACES), Fall 1997-Spring 1998, Fall 2000-Spring 2001, children receiving English language assessments both times.

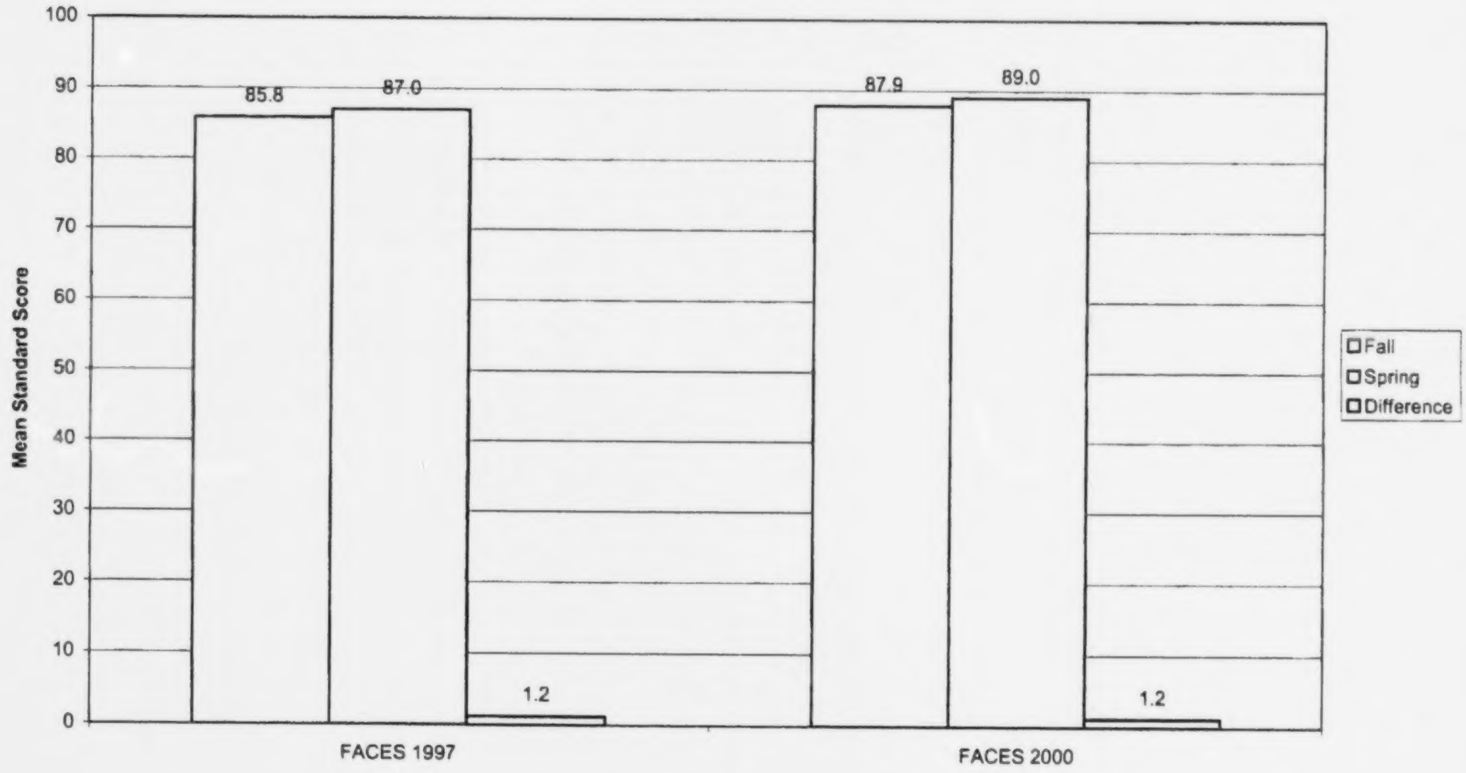
While this finding indicates that children were making more progress in Head Start in the important pre-reading skill of letter recognition, the result shows that Head Start programs still have a way to go in bringing children closer to or even up to the national norm. The Early Childhood Longitudinal Study of the Kindergarten class of 1998 found that a majority of U.S. children knew the letters of the alphabet upon entering kindergarten (Zill & West, 2001). Preliminary findings from randomized intervention studies conducted in Head Start programs in New York state as part of the Head Start Quality Research Consortium studies suggest that children in Head Start can make strikingly larger gains in letter recognition and related skills with certain research-based, literacy-focused curricula (Fischel, Storch, Spira, & Stolz, 2003).

Head Start children in FACES 2000 demonstrated greater gains in knowledge of book and print conventions than they had in FACES 1997. In both studies, when Head Start children were asked, they could show the assessor the front of a storybook and open it to where the adult should start reading. But children in FACES 1997 showed no advance in this sort of book knowledge between the fall and the spring. Mean scores on this task were 1.83 in the fall and 1.85 in the spring. By contrast, in FACES 2000, mean scores went from 1.61 in the fall to 2.46 in the spring. This was a statistically significant gain ( $p < .0001$), with the raw score change equal to two thirds of a standard deviation. The difference in fall-spring gains from FACES 1997 to FACES 2000 was statistically significant ($p < .0001$).

### Little Gain in Early Math Skills

Children in FACES 2000 showed very slight gains in early math skills with respect to national averages. Their mean standard scores on the WJ-R Applied Problems task went from 87.9 in the fall to 89.0 in the spring. While the gain of 1.2 standard score points was statistically significant ($p < .05$), it amounted to only .08 of a standard deviation. The difference in fall-spring gains from FACES 1997 to FACES 2000 was not statistically significant ($p = .87$) (Figure 1.6).

Figure 1.6. Early Math Standard Scores of Children in Fall and Spring of Head Start Year: FACES 1997 Versus FACES 2000



SOURCE: Head Start Family and Child Experiences Survey (FACES), Fall 1997-Spring 1998, Fall 2000-Spring 2001, children receiving English language assessments both times.

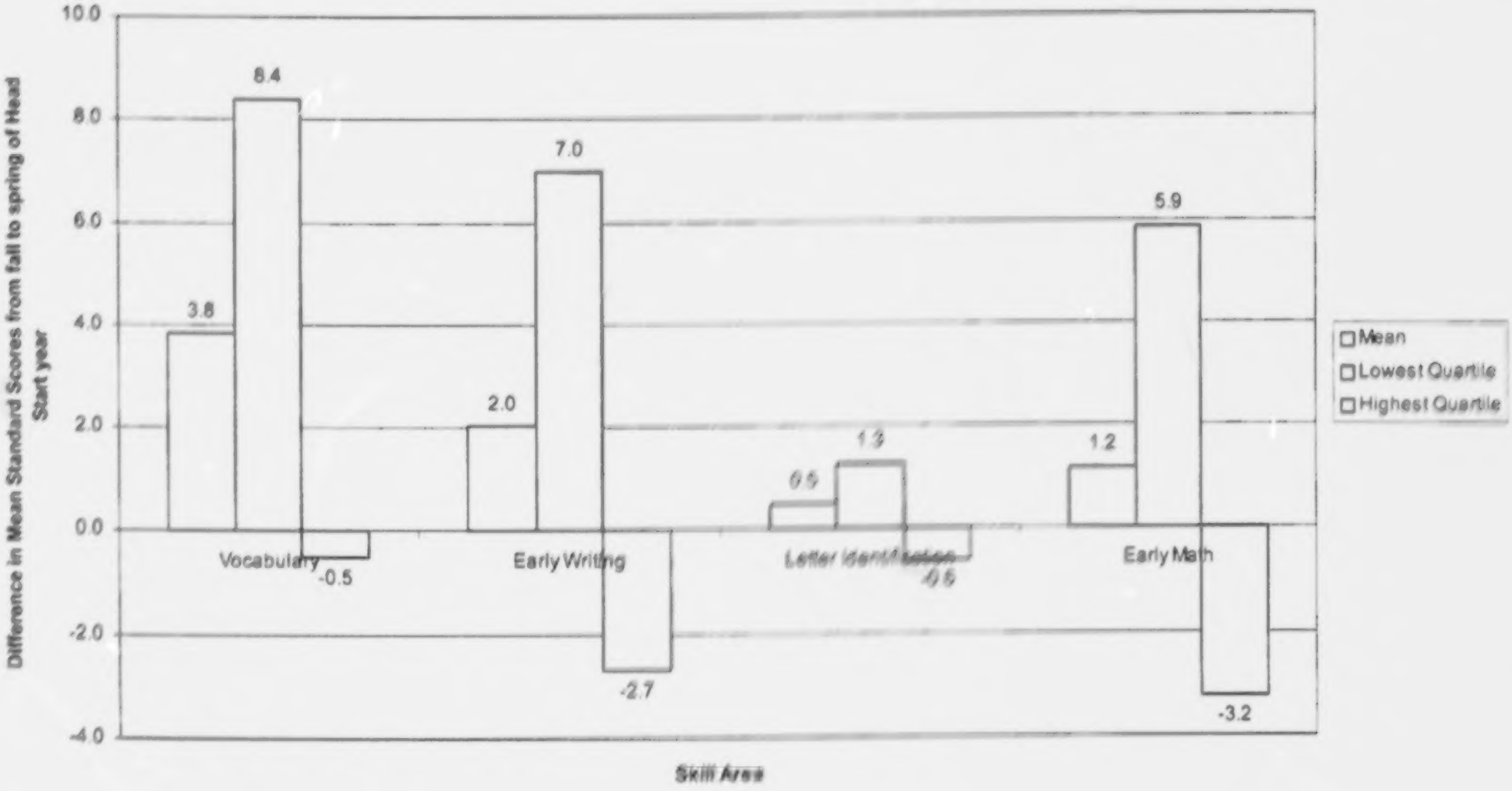
## Greater Gains for Those Who Entered With Lower Skills

Like the 1997 study, FACES 2000 found that children who came to Head Start with lower early literacy and math skills made greater gains in the program than those who came with average skills (Figure 1.7). For example, whereas the average gain in vocabulary standard scores was 3.8 points, the mean gain for those in the lowest quarter of the distribution was 8.4 points. This amounted to more than half of a standard deviation. By contrast, children who were in the highest quarter of the distribution in the fall showed no gain (-0.5) in standard scores by spring. (Differences between mean gain in lowest quartile and overall mean gain, in highest quarter and overall mean gain, and between mean gains in lowest and highest quartiles were all statistically significant, $p < .001$.) However, this finding may be related in part to the tendency of scores to move close to the population mean over successive assessments.

FACES 2000 found a similar picture with respect to changes in early writing and math skills (Figure 1.7). With respect to letter recognition, there was evidence of a slight but significant increase in standard scores (1.26, or .08 of a standard deviation) for children in the lowest quartile in the fall. In FACES 1997, the lowest quartile group had not shown even this relatively slight gain in standard scores.

Despite the greater gains shown by children who entered in the lowest quarter of the Head Start population, these children ended the year with skills that were still well below average.

**Figure 1.7. Children Who Enter Head Start With Lower Skills Show Larger Gains in Comparison to National Norms**



SOURCE: Head Start Family and Child Experiences Survey (FACES), fall 2000 and spring 2001, children receiving assessments in English both times.

## C. Emergent Literacy Gains of Language-Minority Children

In FACES 2000, children whose English-language skills were not sufficient to enable them to be assessed fully in English, and whose parents spoke Spanish at home, were given a Spanish-language version of the FACES battery in the fall. In the spring, these children received the full assessment battery in English. However, these English-language learners also received two additional tests that enabled the research team to track their language development and pre-reading skills in both English and Spanish. In the fall, they received two components of the battery – vocabulary and letter identification – in English as well as Spanish. In the spring, they received the same two components in Spanish as well as English. Thus, the levels of proficiency achieved and the gains made in vocabulary and letter identification could be determined in both languages.[8]

The dual assessment procedure made it possible to investigate the following research questions:

1. How do the literacy levels and gains of language-minority children in Head Start compare with those of language-majority children?
2. How do the literacy levels and gains of these children vary across the two languages?

Of course, in comparing children to test norms, a different set of norms has to be used for test performance in Spanish. The Spanish-language knowledge of Head Start children from Spanish-speaking homes was compared to norms based on samples of children drawn in Mexico and Puerto Rico combined (Dunn, Padilla, Lugo, & Dunn, 1986).

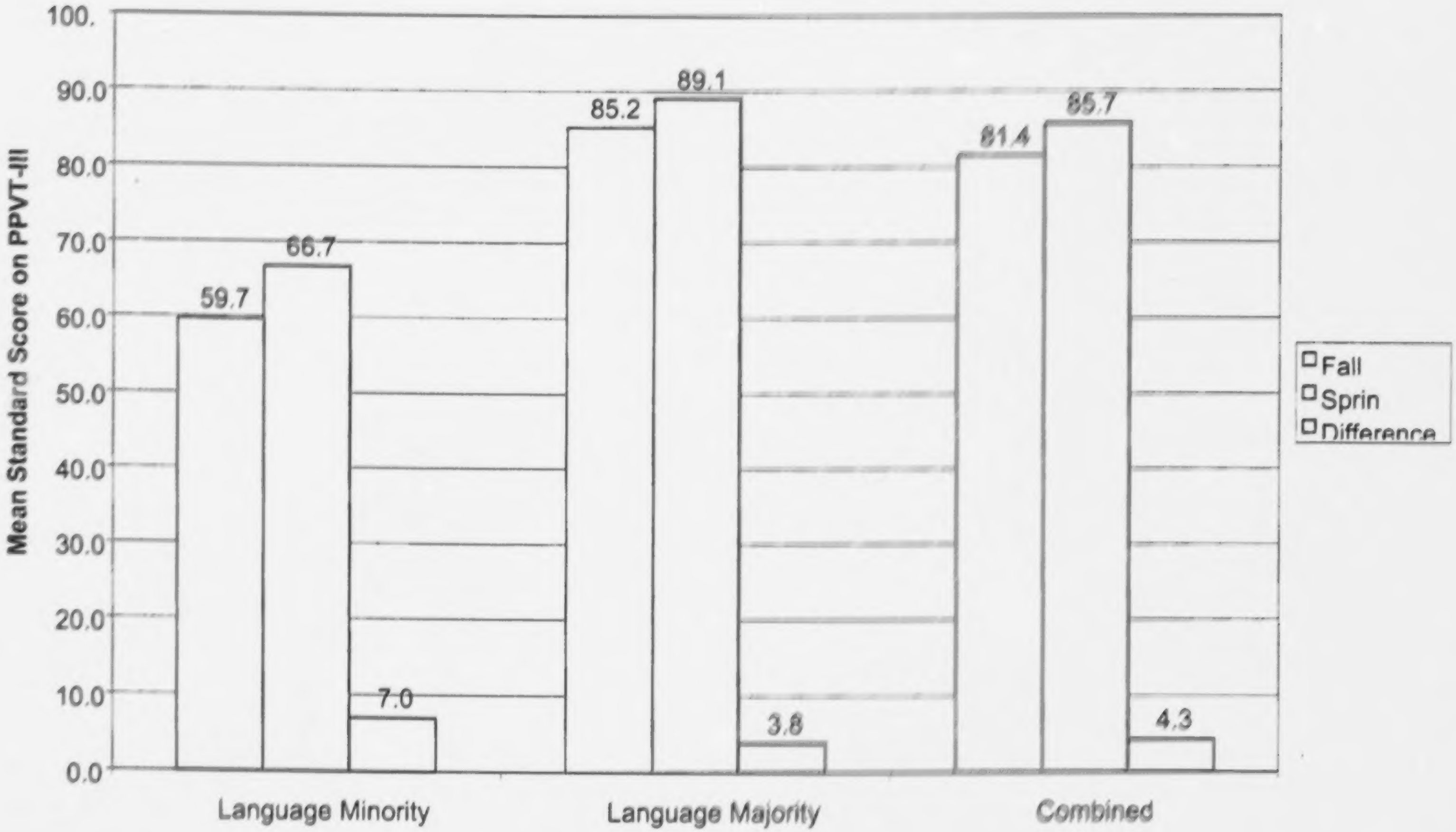
### Knowing the Meaning of English Words

Spanish-speaking children in Head Start entered with English vocabulary skills considerably behind those of children who came from households where English was the primary language spoken in the home. They made greater gains over the course of the Head Start year. But their English vocabulary knowledge remained behind that of other children at the end of the year. The mean standard score on the PPVT-III for Spanish-speaking children in fall 2000 was 59.7, two standard deviations below the national norm. By the spring, their mean score had risen to 66.7, a gain of 7 points. But this was still more than 20 points (or one and a third standard deviations) lower than the mean score in vocabulary for language-majority children in Head Start (Figure 1.8).

---

[8] English-language learners whose home language was something other than Spanish were not given any direct assessment in their native languages. They received only the full assessment battery in English in the spring. For further details, see the Appendix.

Figure 1.8. English-Language Vocabulary Skills of Head Start Children: Spanish-Language Minority, Language Majority, and Combined Population



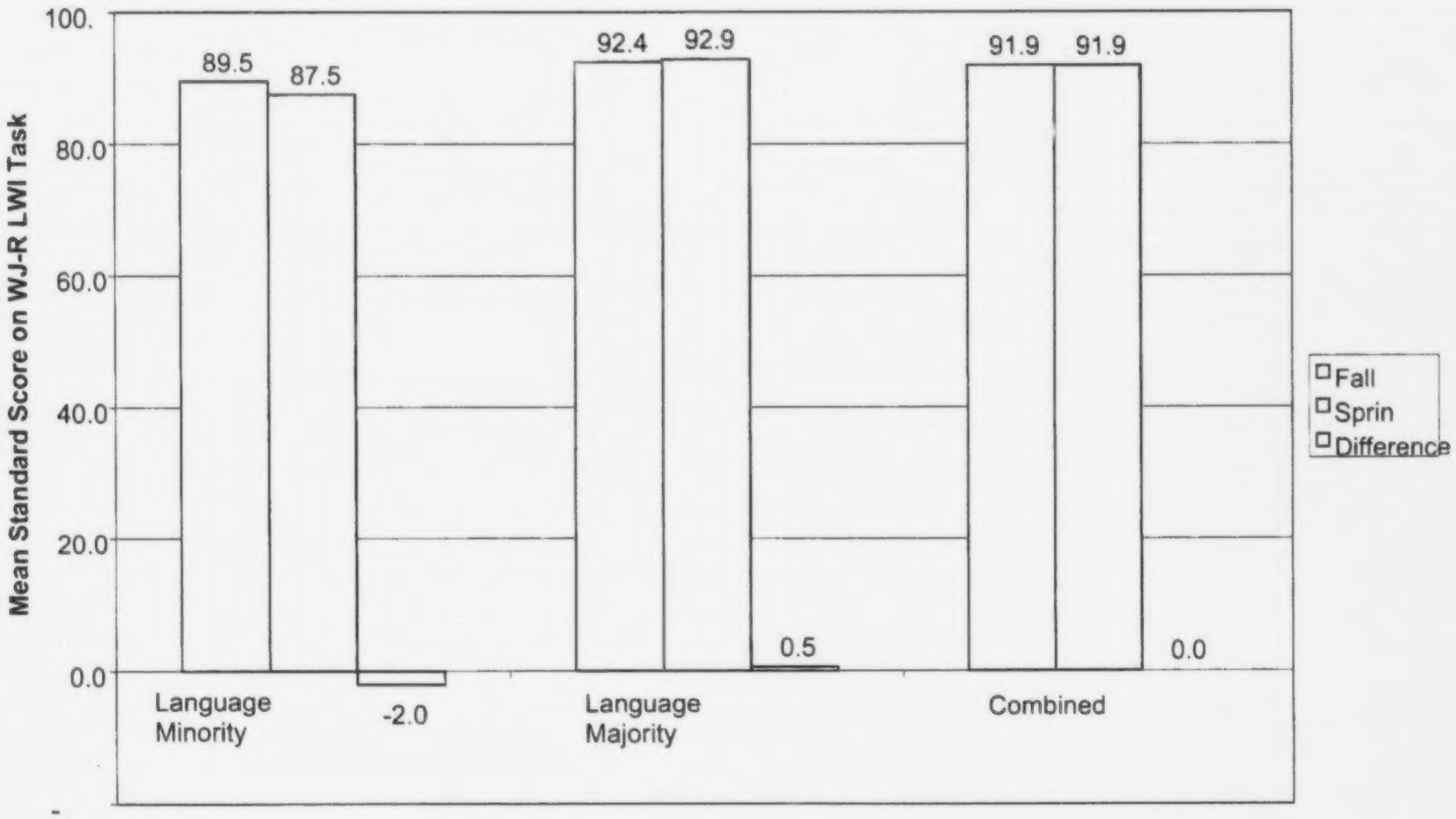
When the vocabulary scores of language-minority and language-majority children were added together, the combined Head Start student population began the year with a mean standard score of 81.4, and ended the year with a mean standard score of 85.7. This represented a gain of 4.3 standard score points.

### Identifying Letters in English

Spanish-speaking children in Head Start entered with English letter identification skills that were only slightly behind those of language-majority children. However, they did not make gains in these skills over the course of the year, at least not in comparison to national averages. The mean standard score of Spanish-speaking children tested in English on the WJ-R Letter-Word Identification test for Spanish-speaking children in fall 2000 was 89.5, about two thirds of a standard deviation below the national norm. By the spring, their mean score was 87.5, which was not significantly different from the fall score (Figure 1.9).

When the letter identification scores of language-minority and language-majority children were added together, the combined Head Start student population began and ended the Head Start year with a mean standard score of 91.9.

Figure 1.9 Letter-Identification in English By Head Start Children: Spanish-Speaking Language Minority, Language Majority, and Combined Population

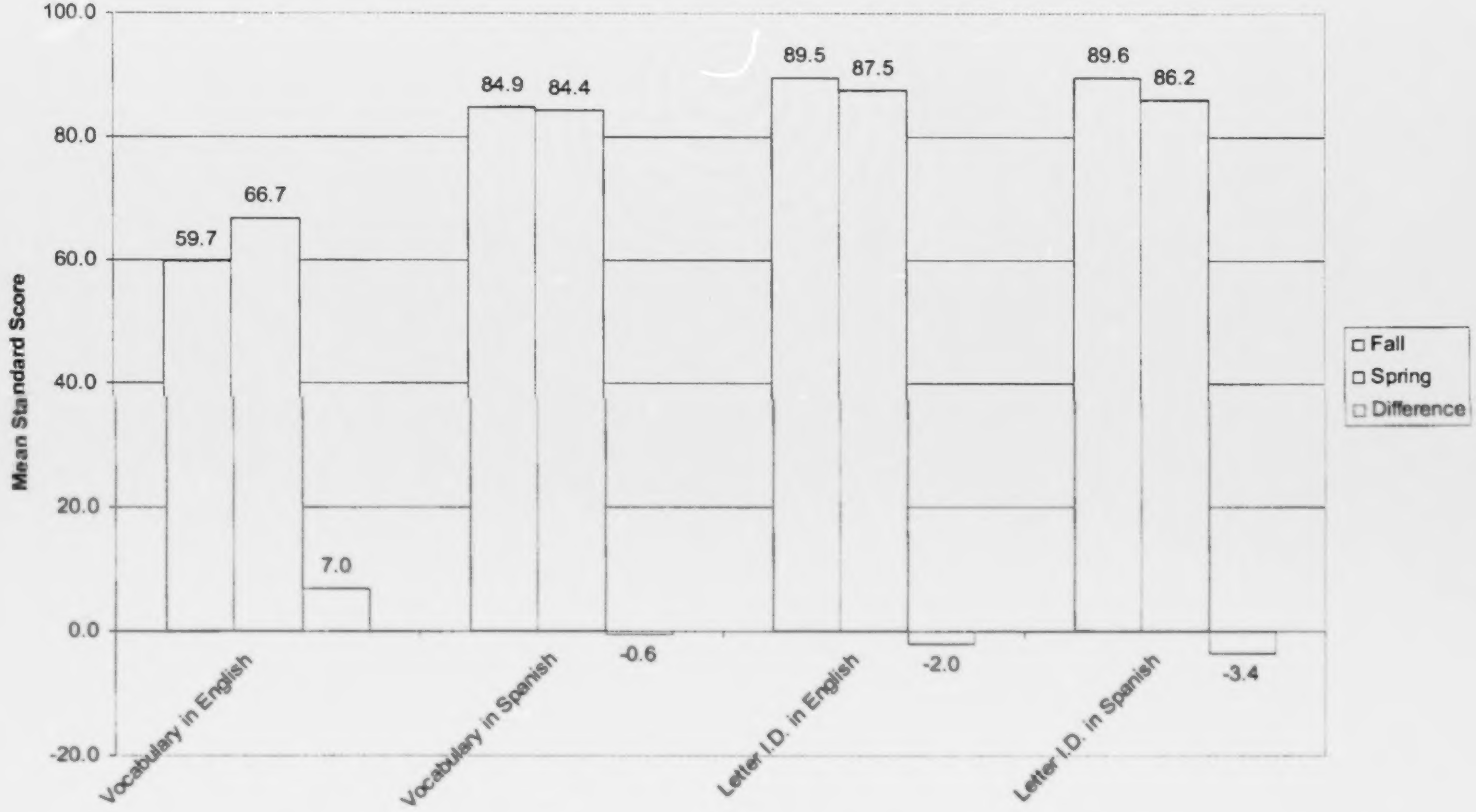


## Changes in English Language Skills Versus Changes in Spanish Skills

The Spanish vocabulary knowledge of Spanish-speaking Head Start children was quite comparable, in standard score terms, to the knowledge levels of English-speaking Head Start children in English. While the Spanish children made gains in English vocabulary knowledge during the year, they did not lose ground in Spanish vocabulary knowledge. The mean standard score for Spanish vocabulary knowledge was 84.9 in the fall and 84.4 in the spring (Figure 1.10).

The ability of Spanish-speaking Head Start children to identify letters of the alphabet in Spanish was comparable, in standard score terms, to the ability of English-speaking Head Start children to identify letters in English. It was also comparable to their own ability to identify letters in English. However, the Spanish-speaking children showed no gains against norms in their ability to identify letters in either language. The mean standard score for Spanish letter identification was 89.6 in the fall and 86.2 in the spring (Figure 1.10).

Figure 10. Vocabulary and Letter Identification Skills in English and in Spanish of Head Start Children From Spanish-Speaking Language Minority Families

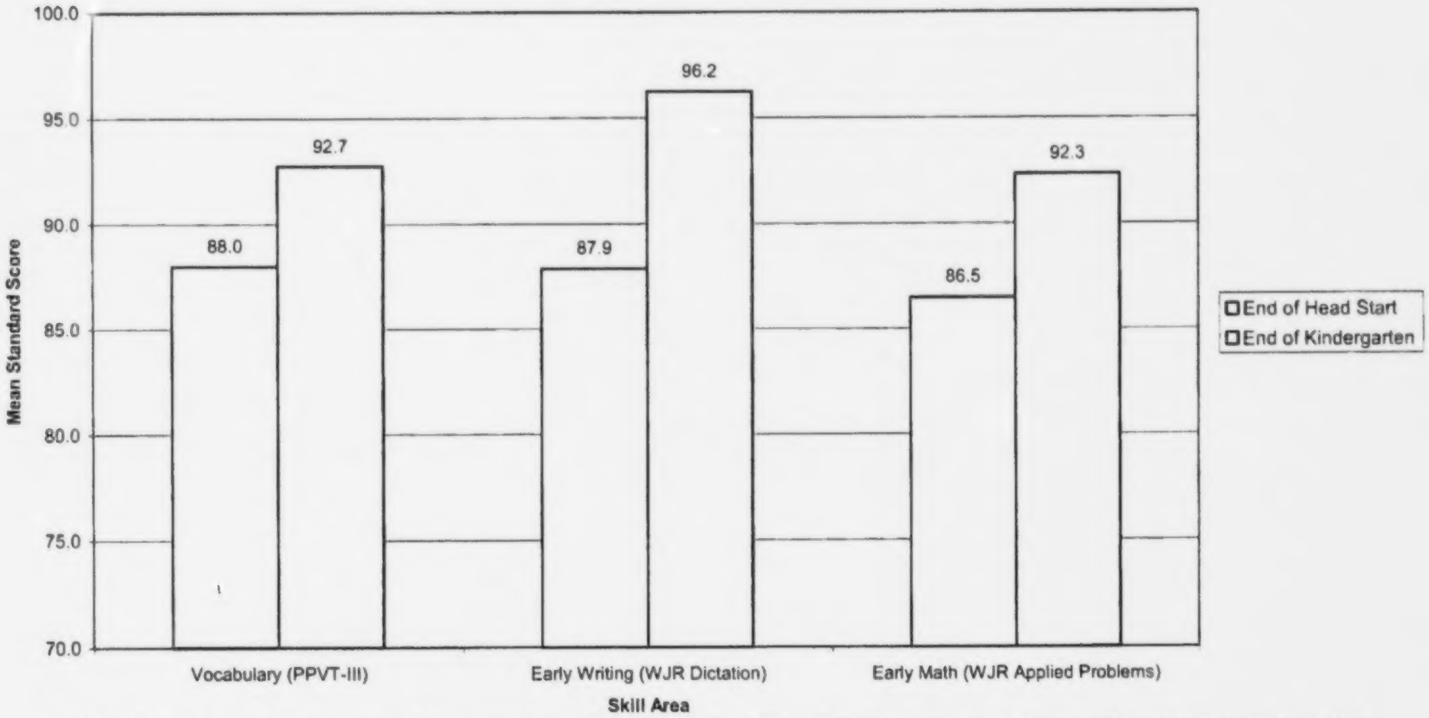


## D. Growth of Knowledge and Skills in Kindergarten

Assessment data are not yet available from FACES 2000 on the knowledge and skills of Head Start graduates by the time they reach the end of kindergarten. But analyses show something about the kindergarten experiences of Head Start graduates from longitudinal follow-ups to FACES 1997. For example, whereas Head Start children in 1997-1998 could identify about one third of the letters of the alphabet by the end of Head Start, by the end of kindergarten 82 percent could identify most or all letters. More than half – 52 percent – could recognize beginning sounds of words by the end of kindergarten.

Analyses from assessments also show in follow-ups to FACES 1997 that Head Start children continue to make advances against national norms during their first year in elementary school. For example, vocabulary standard scores went from 88.0 at the end of Head Start to 92.7 at the end of kindergarten. This was an increase of 4.7 points, or nearly one third of a standard deviation. Similarly, early math scores went from a mean standard score of 86.5 at the end of Head Start to a mean of 92.3 at the end of kindergarten. This was a gain of 5.8 points, or nearly 40 percent of a standard deviation. Early writing scores showed a gain in kindergarten of more than half a standard deviation. They went from 86.5 at the end of Head Start to 92.3 at the end of kindergarten, an increase of 8.3 points (Figure 1.11). Similar analyses will be accomplished on follow-ups to the FACES 2000 sample, once data collection is completed in Spring 2003.

Figure 1.11. Head Start Graduates Show Further Progress Toward National Norms in Kindergarten



SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 and 1999 (combined) and Spring 1999 and 2000 (combined). Ns = 1215-1252.

Another important finding from the follow-up studies to FACES 1997 was that children's cognitive scores in Head Start were predictive of their achievement by the end of kindergarten. Not only were children's scores at the end of Head Start predictive, the size of the *gains* they made in Head Start from the skill levels at which they entered the program, were predictive of later achievement. The evidence for these conclusions is presented in Chapter VII of this report. These conclusions support the predictive validity of the cognitive assessments carried out in FACES.

## CONCLUSIONS

Cognitive assessment data from FACES 2000 showed that most children entered Head Start with early literacy and math skills well below national averages. The typical Head Start child was found to enter at about the 16th percentile in vocabulary and early writing skills, at about the 31st percentile in letter recognition and at about the 21st percentile in early math, when compared to the full spectrum of American children in the same age range. There was considerable diversity in skill levels among Head Start children, however. The highest quarter of Head Start children were at or above the national average (50th percentile) in early language and number skills, while the lowest quarter of children ranked in the lowest 2 percent of all U.S. preschoolers in these areas.

Children made gains toward national averages during the Head Start year, especially with respect to vocabulary knowledge and early writing skills. Children showed greater progress in letter recognition skills than they had in 1997-1998, but they still did not advance as much as the typical child in this area. Nor did they make gains toward national averages with respect to early math skills. Children who entered the program with lower levels of knowledge and skill showed larger gains during the program year, yet still lagged considerably behind national averages. Children who started with higher assessment scores in the fall wound up with higher scores in the spring, but showed less dramatic gains. Language-minority children in Head Start showed significant gains in English vocabulary skills without declines in their Spanish vocabulary skills. They did not gain in letter recognition skills.

Based on follow-up of the 1997-1998 cohort, Head Start graduates showed further progress toward national averages during kindergarten. Gains of between a third to more than half a standard deviation were observed in vocabulary, math, and writing skills during kindergarten. Most Head Start graduates could identify the letters of the alphabet by the end of kindergarten and more than half could recognize beginning sounds of words. Nevertheless, Head Start graduates remained behind their more advantaged peers in early achievement. The size of gains that children made while in Head Start were predictive of their achievement levels by the end of kindergarten. This seems to suggest that efforts to bolster achievement gains in Head Start should result in higher achievement for program graduates in early elementary school.

# REFERENCES

Barnett, W.S. (1998). Long-term effects on cognitive development and school success. Pp. 11-44 in *Early Care and Education for Children in Poverty*, edited by W.S. Barnett & S.S. Boocock. Albany, NY: State University of New York Press.

Dunn, L.M., Padilla, E.R., Lugo, D.E., & Dunn, L.M. (1986). *Test de Vocabulario en Imagenes Peabody.* Circle Pines, MN: American Guidance Service.

Fishel, J.E., Storch, S.A., Spira, E.G., & Stolz, B.M. (2003). *Enhancing emergent literacy skills in Head Start: First year curriculum evaluation results.* Presented at Biennial Meeting of the Society for Research in Child Development, Tampa, FL.

Haskins, R. (1989). Beyond metaphor: The efficacy of early childhood education. *American Psychologist,* 44(2), 274-282.

McKey, R.H., Condelli, L., Ganson, H., Barrett, B.J., McConkey, C., & Plantz, M.C. (1985). *The impact of Head Start on children, families, and communities.* (DHHS Publication No. OHDS 85-31193). Washington, DC: U.S. Government Printing Office.

Phillips, M., Brooks-Gunn, J., Duncan, G.J., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the black-white test score gap. Pp. 103-145 in *The Black-White Test Score Gap*, edited by C. Jencks & M. Phillips, Washington DC: Brookings Institution Press.

Rosenthal, R. & Rosnow, R.L. (1984*). Essentials of behavioral analysis: Methods and data analysis.* New York: McGraw-Hill.

Zill, N., & West, J. (2001). *Entering kindergarten: A portrait of American children when they begin school.* Washington, DC: National Center for Education Statistics, U.S. Department of Education.

# Chapter II: Social Skills and Problem Behavior of Head Start Children and Their Adjustment to Kindergarten

The Head Start Family and Child Experiences Survey (FACES) provides information about desirable and undesirable social behavior that children who attend Head Start display when they enter the program, at the end of the program year, and at the end of the first year of elementary school. The information is obtained through behavior ratings provided by parents, Head Start teachers, and kindergarten teachers. Desirable behavior includes cooperation with adults, friendly play, and caring for and sharing with other children. Undesirable behaviors include disruptive or overly aggressive behavior toward other children or defiant behavior toward adults, hyperactivity, and excessive shyness, lack of self-confidence, and social withdrawal.

Information about children's social behavior is important because children's academic achievement and adjustment once they get to school depend not only on their mastery of basic intellectual skills, but also on their acquisition of social skills and positive approaches to learning. They also depend on the amelioration of negative behavior patterns they may have acquired that are likely to be disruptive or counter-productive in the school setting (Pianta & McCoy, 1997; Zill & West, 2001). Thus, data on behavioral changes that occur during and after program participation is helpful in evaluating how well the Head Start program is performing in preparing children for school, and what changes and reforms may be needed to improve program performance.[9]

## RESEARCH QUESTIONS

Behavior rating data from FACES 2000 were used to address the following research questions:

1. Do children show significant gains in social skills and significant declines in problem behavior during the Head Start year?
2. How do gains vary across behavior areas and among children who enter the program with lower levels of social skills or higher levels of problem behavior?
3. Are the behavioral gains that Head Start children make changing? Did they change significantly between the 1997-1998 program year and the 2000-2001 year?
4. How well do Head Start children graduates adjust to the demands of elementary school? Is their behavior at the end of Head Start predictive of adjustment in kindergarten?

In making these comparisons, data are presented on the full sample of children, including both those who were assessed in English in both the fall and spring of the Head Start year, and those who were assessed primarily in Spanish in the fall and primarily in English in the spring.

---

[9] This chapter focuses on social-emotional measures of school readiness. See Chapter I for information about cognitive measures and the Appendix for a complete listing of measures used in FACES.

# FINDINGS

Children who entered Head Start in the fall of 2000 showed gains in cooperative classroom behavior over the course of the program year. These gains were similar to but significantly larger than those shown by Head Start children in the 1997-1998 program year. Children who entered Head Start with lower levels of social skill made greater gains than those who entered with higher skill levels did. This was also the case in the first round of FACES. According to Head Start teachers, children showed small but significant declines in hyperactive behavior during the program year. They did not show significant declines in overall problem behavior, aggressive behavior, or withdrawn behavior. However, children who entered the program with high levels of problem behavior showed significant improvement, and this was true in all areas of problem behavior. According to Head Start parents, children showed significant improvement in overall problem behavior, aggressive behavior, and hyperactive behavior during the program year. Those who displayed high levels of problem behavior at the start of the year showed greater improvement than those who entered with average or lower levels of undesirable behavior did.

## A. Cooperative Classroom Behavior of Head Start Children

Lead teachers in the FACES 2000 national sample of Head Start classes were asked to rate the cooperative classroom behavior of children in their classes. These ratings were obtained early in the program year (October-November) and toward the end of the year (April-June). The scale was partly composed of items dealing with helpful and compliant behavior like, "Follows the teacher's directions," "Joins an activity or group without being told to do so," and "Helps in putting work materials or center property away." Other items dealt with the child's maturity and skill in interacting with other children. Examples include, "Invites others to join in activities," "Waits her or his turn in games or other activities," and "Accepts classmates' ideas for sharing and playing." For each item, the teacher was asked to judge whether the child behaved in the indicated way, "never," "sometimes," or "very often." There were 12 rating items in all, and the scale score could range from zero (all items marked "never") to 24 (all items marked "very often").

### Head Start Children Show Gain in Cooperative Behavior From Fall to Spring

The mean score on the cooperative classroom behavior scale for children in the fall of 2000 was 14.6. This meant that the typical Head Start child engaged in most of the specified forms of cooperative behavior at least "sometimes," and a few forms, "very often." By the spring of 2001, the mean score on the scale had risen to 16.6. This was a statistically significant gain of two scale score points ($p < .001$), or 41 percent of a standard deviation. It meant that the typical child was engaging in two additional forms of cooperative behavior "very often." The fall-spring gain was slightly larger (by .56 points, $p = .0001$) than the comparable gain (1.41 points, $p<.001$) found among Head Start children in the 1997-98 program year (Figure 2.1).

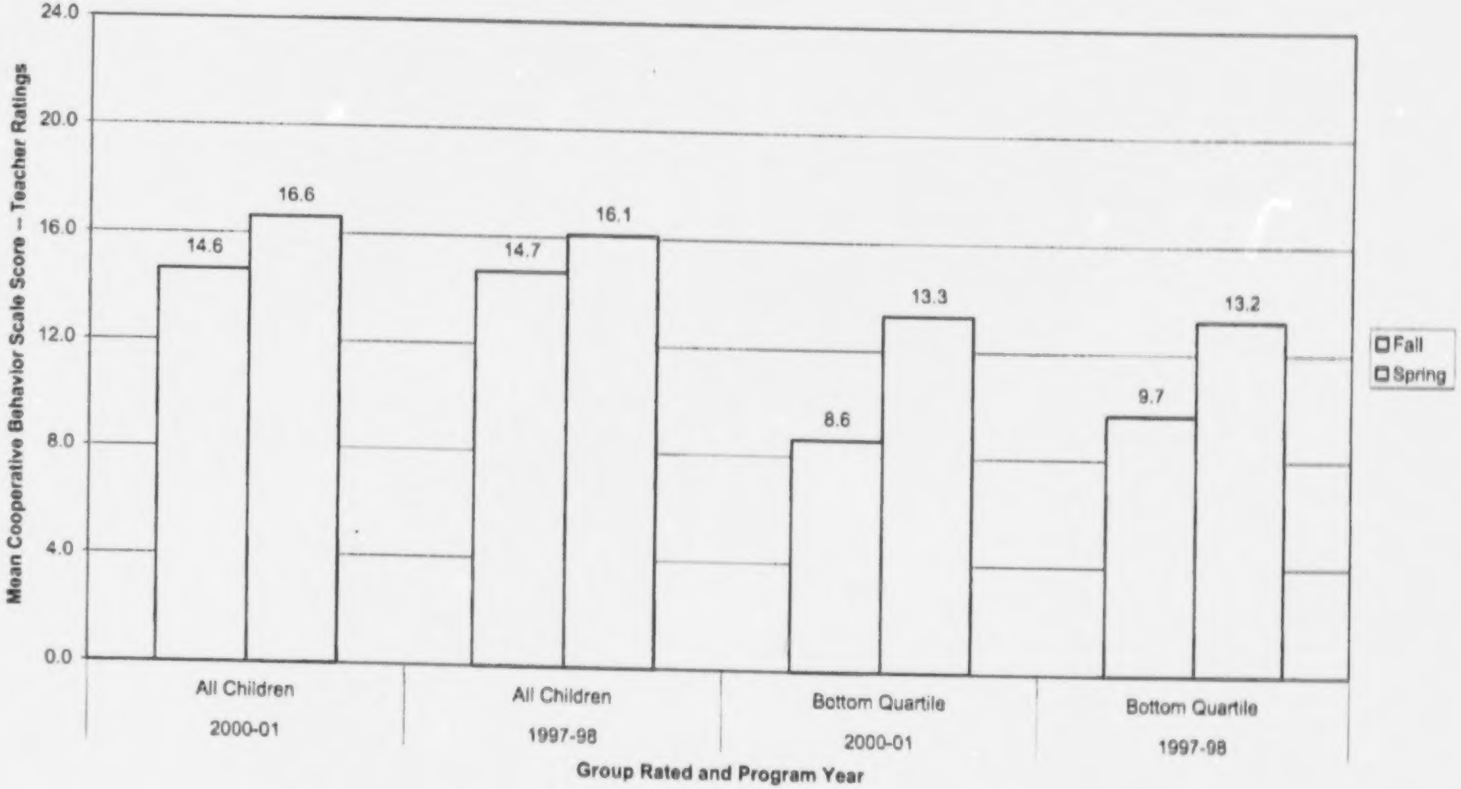
## Children Who Show Less Cooperative Behavior Initially Make Larger Gains

There was considerable diversity in the cooperative behavior levels with which Head Start children entered Head Start. Children in the lowest quartile of the scale score distribution, for example, had a mean score in the fall of 8.6. This implied that these children "never" engaged in several types of cooperative behavior. Children in the highest quartile of the scale score distribution, on the other hand, had a mean score in the fall of 20.5. This implied that, even in the fall, those children engaged in most of the specified forms of cooperative behavior "very often." By the spring, the children in the lowest quartile received a mean score on the cooperative behavior scale of 13.3, whereas the children in the highest quartile had a mean score of 19.8. The gain shown by the children in the lowest quartile in the fall -- 4.7 points, or nearly a full standard deviation -- was more than double that shown by the average Head Start child. It was significantly larger than the gain shown by the average child (difference = 2.7, $p < .001$) or by children in the highest quartile of the fall distribution (difference = 5.4, $p < .001$).

Despite the larger gain shown by children in the lowest quartile in the fall, their mean score remained below the overall spring mean. Likewise, children in the highest quartile in the fall had a mean score in the spring that was significantly higher than the overall mean. This despite a lack of progress (in fact, a slight decline) in their scale scores from fall to spring. Thus, while the gaps between these groups narrowed during the program year, they did not completely close.

The larger gain shown by children who entered Head Start with lower levels of cooperative behavior -- and the decline shown by children who entered with higher levels of cooperative behavior -- may be at least partly attributable to the statistical phenomenon of "regression to the mean." Note however, that the gain shown by children in the lowest quartile was substantially larger than the decline shown by children in the highest quartile, whereas if only regression to the mean was operating, one would expect both changes to be of roughly equal magnitude. The reliability of the teacher rating scale was sufficient (see Appendix) to give us confidence that the larger gain shown by children who entered with lower social skills was not merely a statistical artifact. The larger gain may indicate that Head Start is having a "compensatory" influence on children who entered with lower social skills. In the absence of a control group, however, we cannot say whether the larger gain is attributable to participation in the program or represents a growth spurt that would have occurred among these children whether or not they participated in the program.

Figure 2.1. Head Start Students Show Gains in Cooperative Classroom Behavior During Program Year



SOURCE: Head Start Family and Child Experiences Survey (FACES), 2000-01 (n = 2,191), 1997-98 (n = 1,982).

## B. Problem Behavior of Head Start Children

Lead teachers in the FACES 2000 national sample of Head Start classes were asked to rate the undesirable or problem behavior of children in their classes. These ratings were obtained early in the program year (October-November) and toward the end of the year (April-June). The Behavior Problems scale was partly composed of items dealing with aggressive or defiant behavior like, "Hits or fights with others," "Disrupts ongoing activities," and "Has temper tantrums or hot temper." Other items dealt with inattentive or hyperactive behavior. Examples include, "Can't concentrate, can't pay attention for long," and "Is very restless, fidgets all the time, can't sit still." A third set of items dealt with shy, withdrawn, or depressed behavior. Examples of this set include, "Keeps to herself or himself, tends to withdraw," "Lacks confidence in learning new things or trying new activities," and "Often seems unhappy, sad, or depressed." For each item, the teacher was asked to judge whether the behavioral description was "not true," "somewhat or sometimes true," or "very true or often true" of the child. There were 14 rating items in all and the total scale score could range from zero (all items marked "not true") to 28 (all items marked "very true or often true").

The teacher ratings were also used to calculate three sub-scale scores. The Aggressive Behavior subscale contained 4 items and could range in value from zero to 8. The Hyperactive Behavior subscale was composed of three items and could range in value from zero to 6. And the Withdrawn Behavior subscale contained seven items and could range in value from zero to 14.

### Children Show Reductions in Hyperactive Behavior During Program Year

The mean score on the Total Behavior Problems scale for children in the fall of 2000 was 5.6. This meant that the typical Head Start child engaged in several of the specified forms of aggressive, hyperactive, or withdrawn behavior at least "somewhat or sometimes," or one or two forms of undesirable behavior, "often." By the spring of 2001, the mean score on the scale was 5.2. This was not a statistically significant change (*-.37, p=.071*) (Figure 2.2). The fall and spring levels of problem behavior were similar to those found among Head Start children in the 1997-1998 program year. (The mean in the spring of 1998 was 5.2, a difference of .02, *p = .91.*) The lack of statistically significant change in Total Behavior Problems was also indistinguishable from the change picture found in the earlier cohort (difference in differences = -.08, *p=.57)* (Figure 2.3).

There was, however, a small but statistically significant change in mean scores on the Hyperactive Behavior subscale. These went from a mean of 1.36 in the fall to a mean of 1.20 in the spring. This was a change of -.16 scale points (*p = .008*), or 10 percent of a standard deviation. There was a similar decline in Hyperactive Behavior noted in the 1997-1998 program year. (The difference in differences was -.003, *p= .94.*)

Mean scores on the Aggressive Behavior subscale did not change significantly from (mean = 1.76) to spring (mean = 1.71) of the program year (difference = -0.05, *p=.492*). An apparent decline in mean scores on the Withdrawn Behavior subscale from fall (mean = 2.50) to spring (mean = 2.32) was of only marginal statistical significance (difference = -0.18, *p=.06*).

The subscale scores and change patterns were similar to those found in the 1997-1998 program year.

### Children Who Display More Problem Behavior Initially Show Improvement

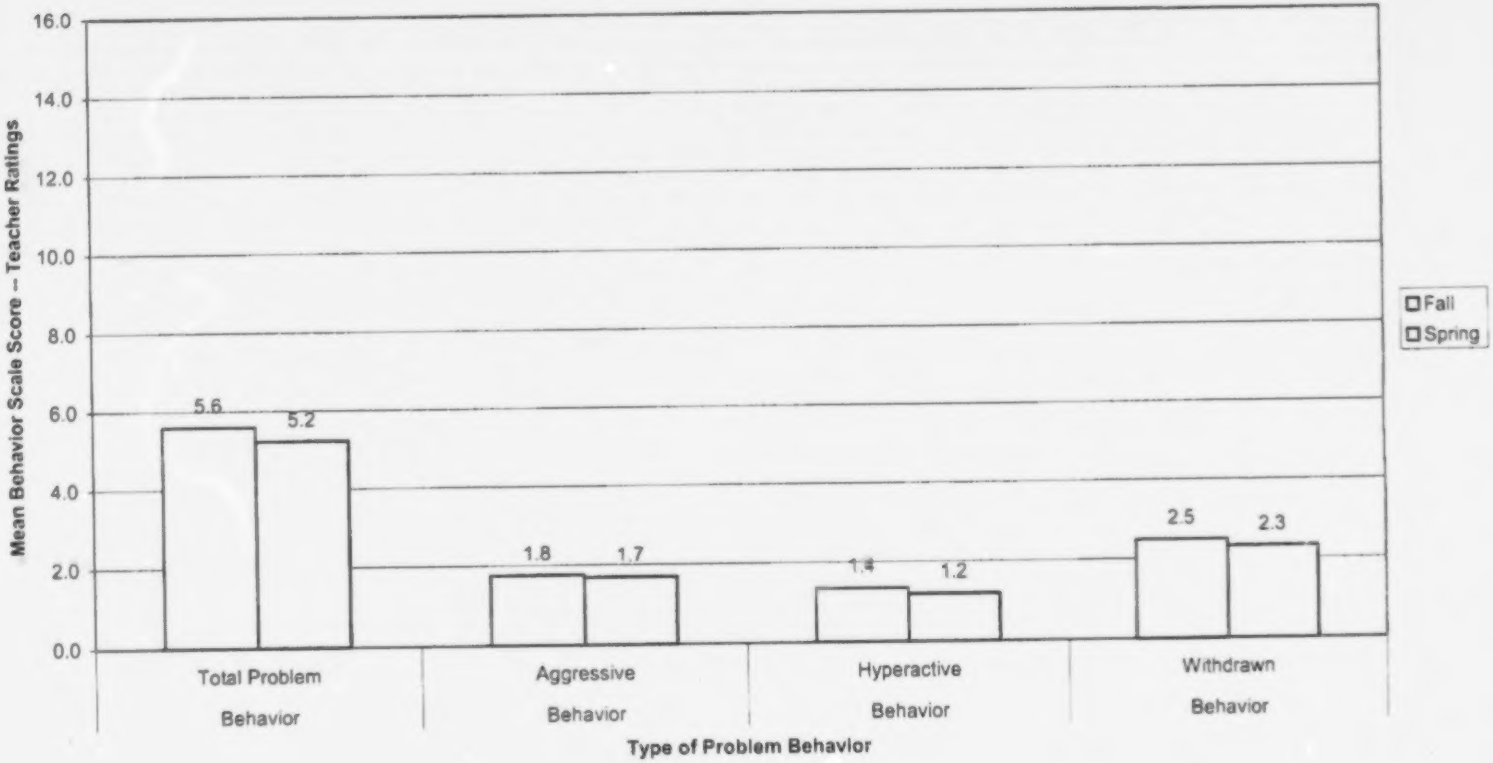
There was substantial diversity in the frequency and severity of problem behavior that children displayed on entry to Head Start. For example, children in the highest quartile of the Total Behavior Problems scale score distribution, for example, had a mean score in the fall of 11.9. This meant that these children displayed nearly a dozen of the specific types of problem behavior covered in the scale at least "somewhat or sometimes." Or they displayed about half a dozen forms of such behavior "often." By contrast, children in the lowest quartile of the scale score distribution had a mean score in the fall of 0.80. Similar diversity was found for the Aggressive, Hyperactive, and Withdrawn Behavior subscales.

Children who entered the program with relatively high levels of problem behavior showed significant declines in their behavior scores from fall to spring. For children in the highest quartile in the fall, the mean score on the Total Behavior Problems scale declined by more than three points (-3.23. $p<.001$), or two thirds of a standard deviation. The mean score on the Withdrawn Behavior subscale declined by nearly two points (-1.88, $p<.001$), while mean scores on the Aggressive (-1.02, $p<.001$) and Hyperactive (-0.94, $p<.001$) subscales declined by about one point each (Figure 2.4). On the other hand, children in the lowest quartile of the problem behavior distribution in the fall showed slight *increases* in their mean scores from fall to spring. For example, children in the lowest quartile on the Total Behavior Problems scale showed an increase of 1.53 points from fall to spring ($p<.001$).

Despite the larger declines shown by children in the highest quartile in the fall, their mean scores remained above overall spring means. Likewise, children in the lowest quartile in the fall had mean scores in the spring that were significantly lower than the overall means. This despite the slight increases they showed in rated problem behavior. Thus, as with cooperative behavior, gaps between high, average, and low problem behavior groups narrowed during the program year, but did not completely close.

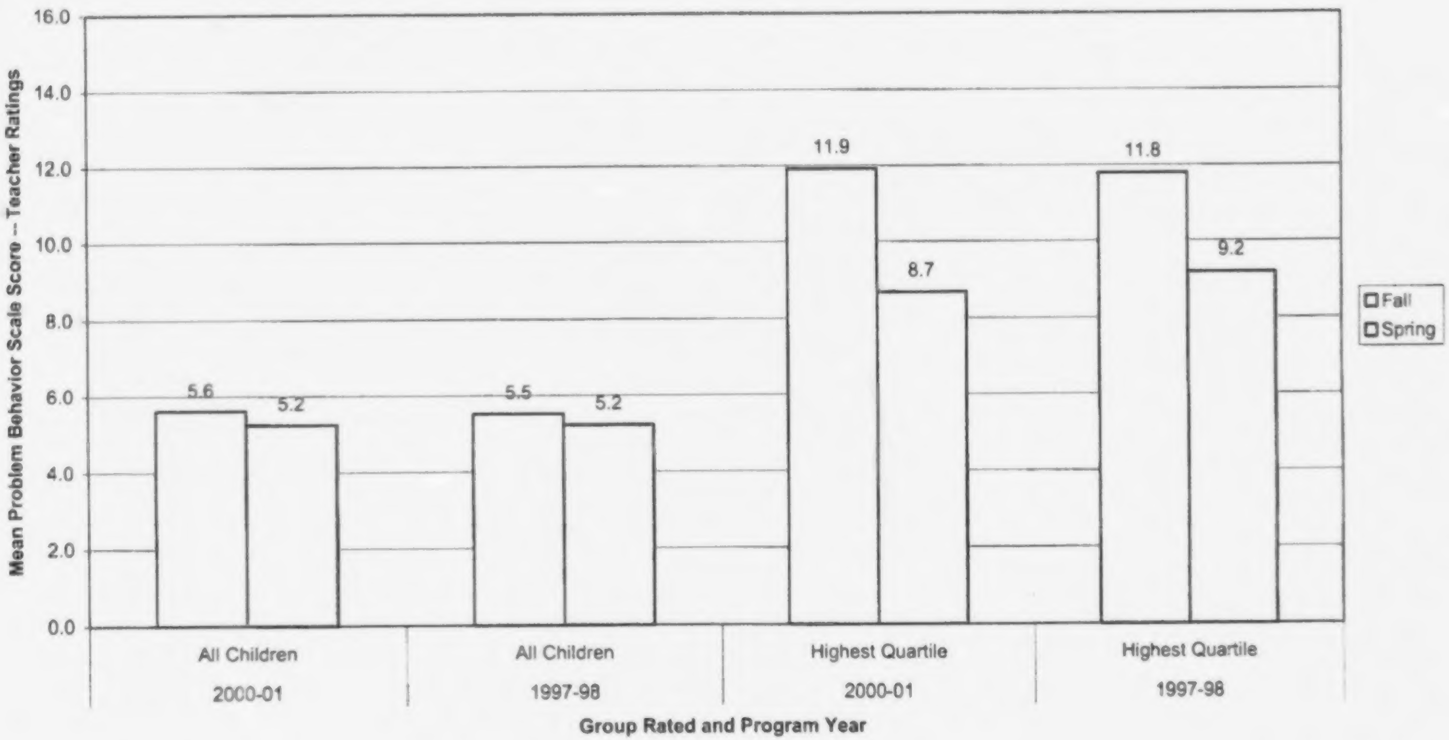
The significant declines in problem behavior shown by children who enter with higher levels of such behavior may indicate that Head Start is having a "compensatory" influence on these children, as the program was intended to have. In the absence of a control group, however, we cannot say whether the larger declines are attributable to participation in the program or represent maturation processes that would have occurred among these children whether or not they participated in the program. Also, the changes that were found in the high and low behavior problem groups may at least partly reflect measurement error and "regression to the mean," as noted in connection with the cooperative behavior ratings. Once again, however, we note that the changes were much larger in the group that showed initial difficulties than in the group that was relatively problem-free at program entry.

Figure 2.2. Head Start Students Show Significant Reductions in Hyperactive Behavior During Program Year



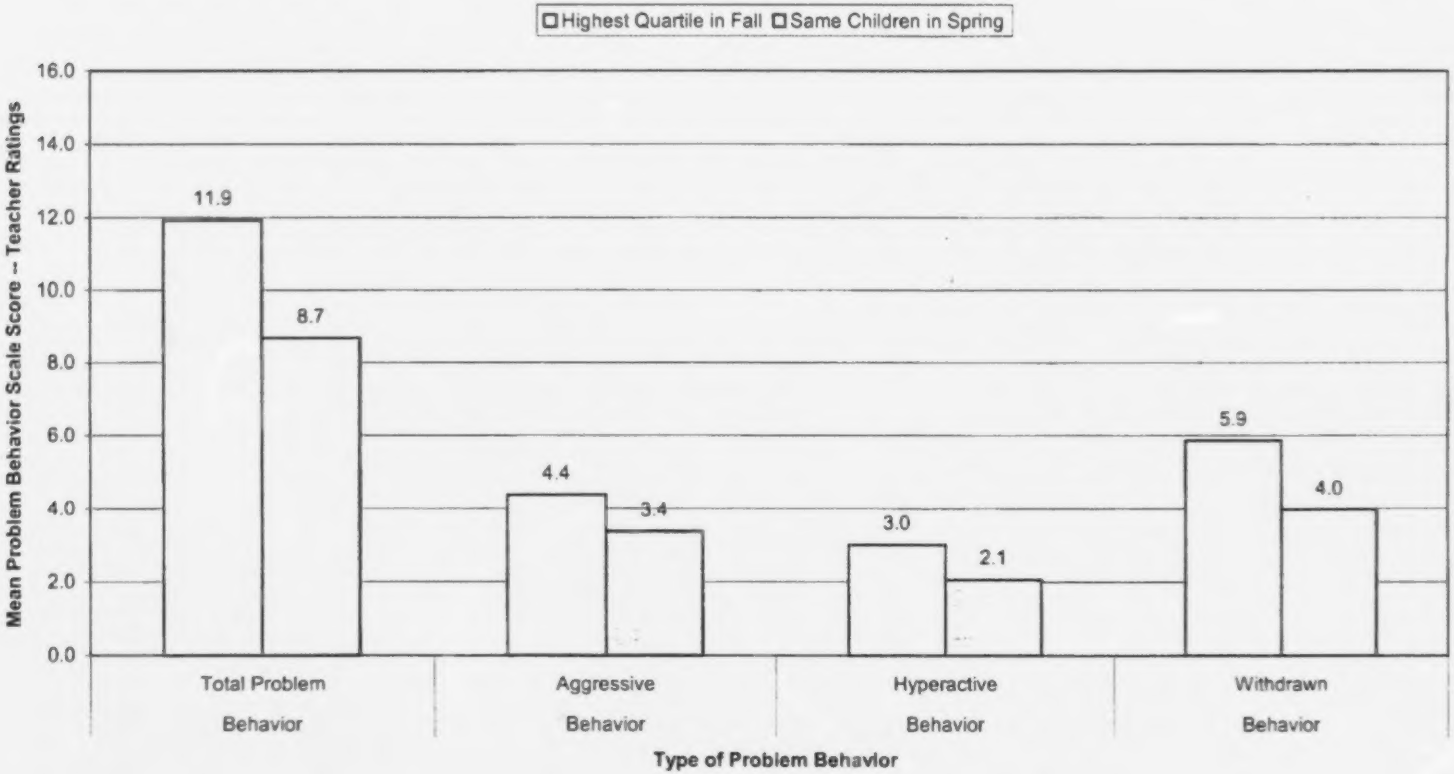
SOURCE: Head Start Family and Child Experiences Survey (FACES), 2000-2001, (n = 2,073-2,130).

**Figure 2.3. Head Start Students Showed Similar Patterns of Problem Behavior in 2000-2001 and 1997-1998 Program Years**



SOURCE: Head Start Family and Child Experiences Survey (FACES), 2000-01 (n = 2,130), 1997-98 (n = 1,953).

**Figure 2.4. Head Start Students Who Enter With High Levels of Problem Behavior Show Improvement During Program Year**



SOURCE: Head Start Family and Child Experiences Survey (FACES), 2000-2001, (n = 2,073-2,130).

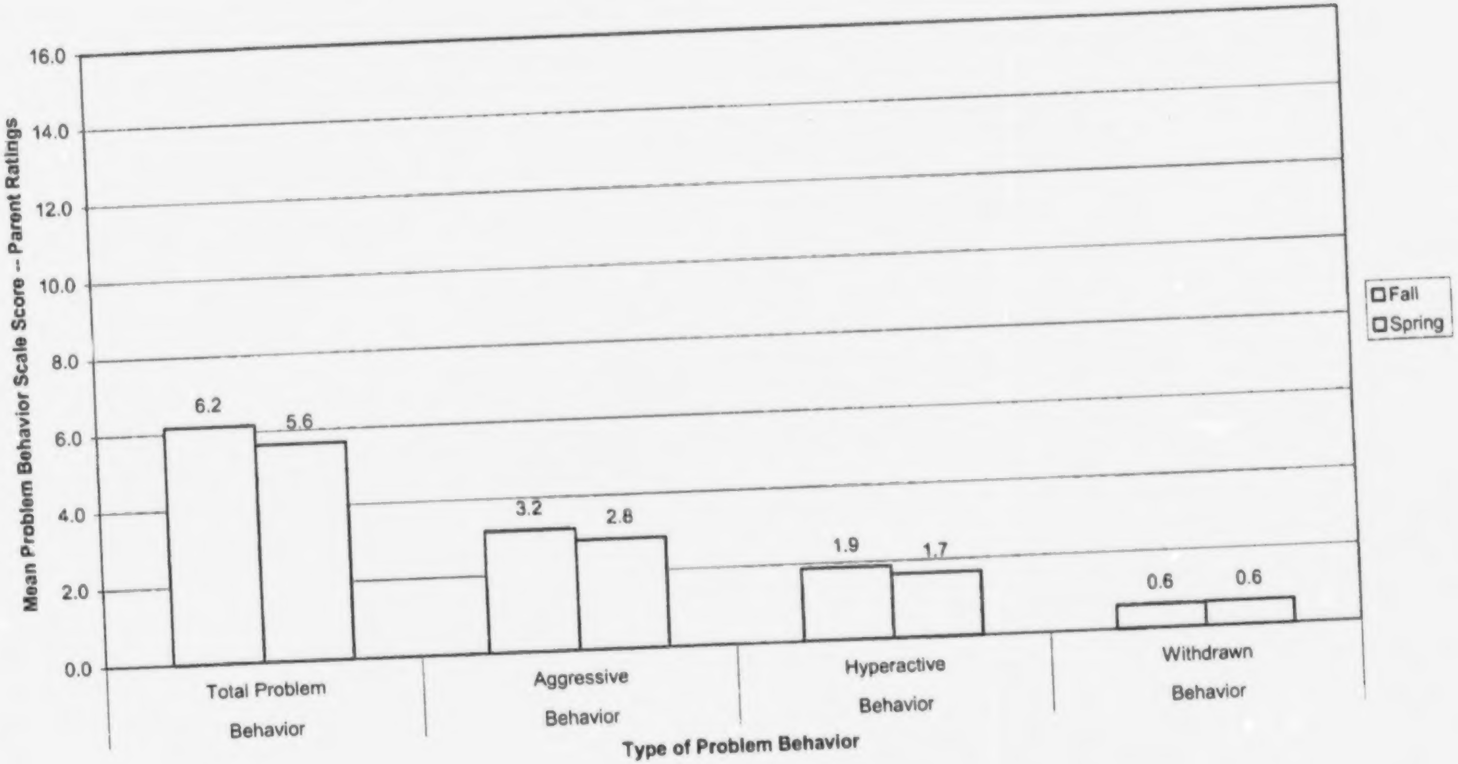
### Parents Report Improvements in Children's Problem Behavior During Head Start

Parents of children studied in FACES 2000 were asked to rate the problem behavior of their children, using scales similar to those completed by teachers. The fall and spring behavior ratings completed by parents showed significant improvements in overall problem behavior. They showed improvements in the Aggressive as well as the Hyperactive subscale. Significant declines were found not just among children who entered the program with high levels of problem behavior, but in overall mean scores as well (Figure 2.5).

Scores on the Total Behavior Problem scale completed by parents could range from zero to 24. Mean scores went from 6.17 in the fall to 5.64 in the spring. The decline of 0.52 points was statistically significant ($p < .001$) and represented a change of 14 percent of a standard deviation. Scores on the Aggressive Behavior subscale could range from zero to 8. Mean scores went from 3.15 in the fall to 2.85 in the spring (difference = -0.31, $p < .001$). Scores on the Hyperactive Behavior subscale could range from zero to 6. Mean scores went from 1.87 in the fall to 1.66 in the spring (difference = -0.22, $p < .001$). Scores on the Withdrawn Behavior subscale could range from zero to 6. Mean scores were 0.61 in the fall and 0.63 in the spring (difference = 0.02, $p = .539$).

Parent behavior ratings in the 2000-2001 FACES sample showed lower levels of problem behavior in their children at the end of the program year than had parents in the 1997-1998 FACES study. This was the case for Total Behavior Problems ($p = .023$), Aggressive Behavior ($p = .025$), and Withdrawn Behavior ($p < .001$). The FACES 2000 parents also reported larger declines in problem behavior from fall to spring than did the FACES 1997 parents. This was the case for Total Behavior Problems ($p < .002$) and Aggressive Behavior ($p < .001$), but not for Hyperactive or Withdrawn Behavior.

Figure 2.5. Parents of Head Start Children Report Reductions in Problem Behavior During Program Year



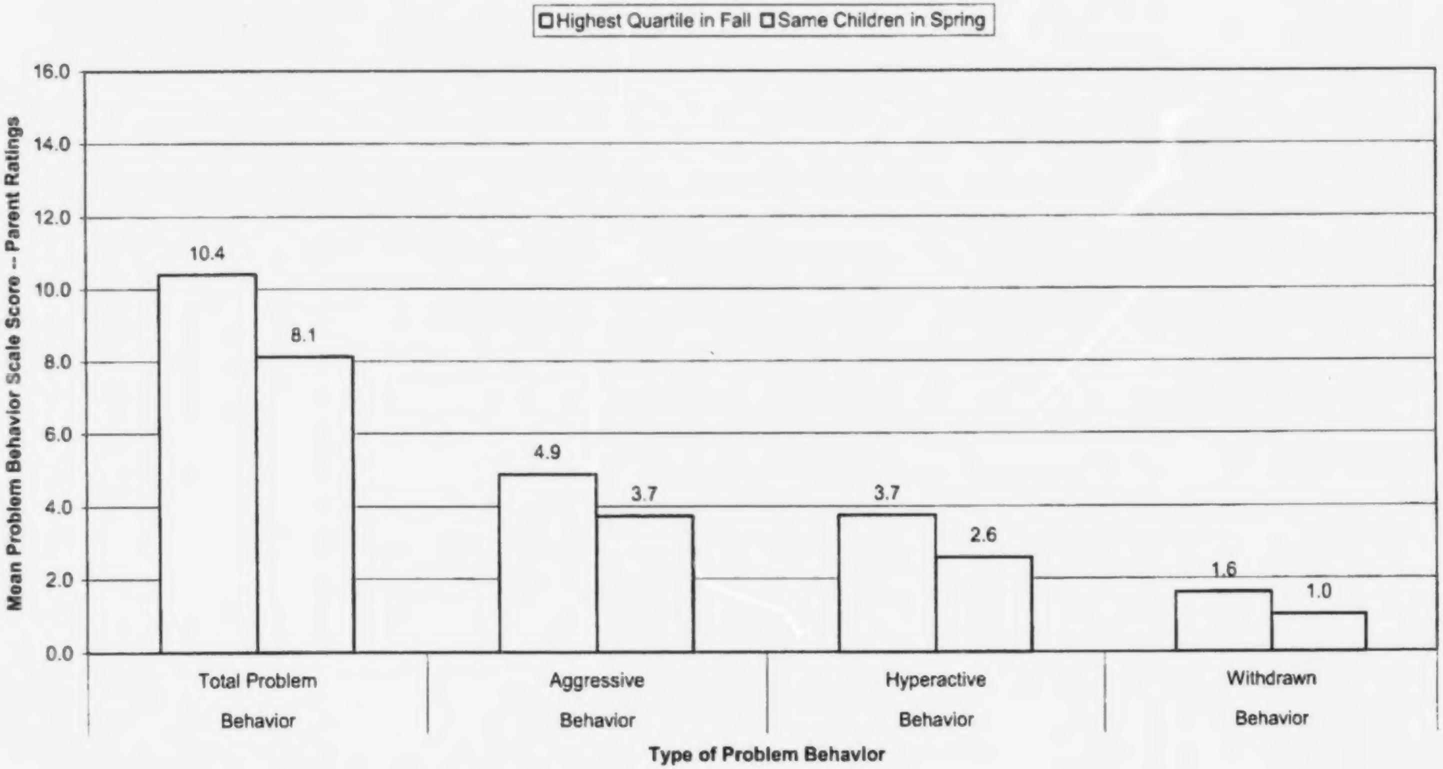
## Parent Ratings Show Greater Improvement for Children With Higher Initial Levels of Problem Behavior

As found with the behavior ratings completed by teachers, parent behavior ratings indicated that children who began Head Start with higher levels of problem behavior showed significant and significantly larger declines in such behavior during the program year than children who entered with average or lower levels of problem behavior. For children in the highest quartile of Total Behavior Problem ratings, mean scores declined more than two points from fall to spring (difference = -2.27, $p < .001$). This was significantly larger than the mean decline (difference of differences = 1.74, $p < .001$) and than the change shown by the lowest quartile children (difference of differences = 3.03, $p < .001$). Children in the highest quartile of Aggressive Behavior or Hyperactive Behavior showed declines of more than one rating point. Children in the highest quartile of Withdrawn Behavior showed declines of 0.61 of a rating point. All of these changes were significantly larger than the corresponding changes for all children or for the lowest quartile of children (all $p < .001$) (Figure 2.6).

The parent ratings of problem behavior produced results that were generally consistent with the results from teacher ratings of problem behavior. Both sets of ratings support the conclusion that children show significant improvement in undesirable behavior, and that this improvement is more pronounced for children who enter Head Start with higher levels of problem behavior. Parents were also asked to rate social skills and positive approaches to learning in their children. Social skills items were similar to those presented to teachers (e.g., "Makes friends easily," "Comforts or helps others," and "Accepts friends' ideas in sharing and playing"). Examples of the approaches to learning items included, "Enjoys learning," "Likes to try new things," and "Shows imagination in work and play." There were seven items in this scale, and scores could range from zero (meaning all the items were rated "not true" of the child) to 14 (meaning all the items were rated "very true or often true" of the child).

On this scale, parent ratings did not show overall change from fall to spring of the program year. Part of the reason for this may be that, even in the fall, parents tended to rate their children as exhibiting most or all of these positive behaviors "very often." Average scores were so high that there was not much room for further improvement. Mean scores were 12.1 in the fall and 12.1 in the spring (difference = .004; $p = .958$). There was significant improvement, however, for those children who received relatively low Social Skills and Positive Approaches to Learning ratings from their parents. Children in the lowest quartile of the distribution in the fall showed a gain of 1.2 scale points, going from 9.98 in the fall to 11.18 in the spring ($p < .001$). This change was significantly greater than that for the overall mean ($p < .001$) and that for the highest quartile (difference = 2.22, $p < .001$). Children in the highest quartile showed a slight decline in their mean scores, going from a perfect mean score of 14.0 in the fall to a near perfect mean score of 12.98 in the spring (difference = -1.02, $p < .001$).

Figure 2.6. Parents Report Greater Improvement in Head Start Students Who Enter With High Levels of Problem Behavior



## C. Behavior and Adjustment in Kindergarten

Teacher ratings data are not yet available from FACES 2000 on the cooperative behavior and problem behavior of Head Start graduates by the time they reach the end of kindergarten. But the data reveal something about the kindergarten experiences of Head Start graduates from longitudinal follow-ups to FACES 1997. Behavioral ratings by kindergarten teachers were obtained for 955 children who participated in Head Start during the 1997-1998 and 1998-1999 program years. (The latter children were those, usually aged three at program entry, who participated in Head Start for two years.) The kindergarten teachers rated the behavior of these children on the same Cooperative Classroom Behavior and Problem Behavior scales used by Head Start teachers.

The picture that these ratings provided was generally consistent with that provided by Head Start teachers at the end of the child's last year in Head Start. For example, the subset of children for whom kindergarten ratings were obtained had a mean score on the Cooperative Behavior scale of 17.7 at the end of Head Start.[10] These same children received a mean score from their kindergarten teachers of 18.8. This implies that the Head Start graduates displayed a majority of the cooperative behaviors "sometimes" or "often." The 1.2 scale-point increase in the mean scores from Head Start to kindergarten was statistically significant ($p < .001$).

The mean score on the Total Behavior Problems scale was 4.27 at the end of Head Start, and 4.57 at the end of kindergarten. The difference in means (0.30) was not statistically significant ($p = .208$). This implies that Head Start graduates in kindergarten displayed between four and five kinds of problem behavior at least "sometimes," or one or two types "often." The frequency of problem behavior was similar at the end of Head Start and kindergarten. The one type of behavior that showed a slight but significant *increase* from the Head Start teacher ratings to the kindergarten teacher ratings was Hyperactive Behavior. Mean scores on this subscale went from 0.88 at the end of Head Start to 1.20 at the end of kindergarten (difference = .31, $p < .001$). Mean scores on the Aggressive Behavior subscale were not significantly different (1.38 in Head Start, 1.30 in kindergarten, difference = -.08, $p = .318$). Nor were means on the Withdrawn Behavior subscale (1.99 in Head Start, 2.05 in kindergarten, difference = .06, $p = .208$).

As shown in greater detail in Chapter VII of this report, behavior ratings by Head Start teachers and parents were predictive of children's behavior and adjustment in kindergarten. When teacher and parent ratings at the end of Head Start were combined in a multiple regression equation, they related moderately well (R = .42) to Cooperative Behavior ratings by kindergarten teachers. Teacher ratings of aggressive behavior and cooperative behavior were the scales that showed the strongest relationship with cooperative behavior in kindergarten. A similar result was obtained with respect to problem behavior in kindergarten. When the Head Start rating scales were put into a regression model, they also related moderately well (R = .49) to Total Behavior Problems as rated by kindergarten teachers. The individual scales or subscales that related most strongly to kindergarten behavior problems were Head Start teachers' ratings of aggressive behavior, withdrawn behavior, and cooperative classroom behavior.

---

[10] The fact that the mean Cooperative Behavior rating was slightly higher for the children for whom kindergarten teacher ratings were obtained than for the overall FACES 1997 sample suggests that the follow-up subsample was somewhat biased toward children with better adjustment.

These findings show the importance of children's behavioral adjustment in preschool as a foreshadowing of how they will adjust and perform in elementary school. They also suggest that the positive changes in behavior that Head Start children display during Head Start may well boost the odds of their doing well in elementary school.

## REFERENCES

Pianta, R.C., & McCoy, S.J. (1997). The first day of school: The predictive validity of early school screening. *Journal of Applied Developmental Psychology, 18*, 1-22.

Zill, N., & West, J. (2001). *Entering kindergarten: A portrait of American children when they begin school.* Washington, DC: National Center for Education Statistics, U.S. Department of Education.

# Chapter III: Relationship Between Curricula and Family, Program, and Classroom Characteristics

There is a lack of substantial evidence about the relative efficacy of various types of standardized curricula available to preschool programs and their relationship to children's school readiness. Accordingly, the Head Start Program Performance Standards require that programs have a curriculum, and delineate the areas that must be covered by it, but do not prescribe one. Programs may use curricula from a variety of sources, develop one of their own, or use a combination of curricula. In 2001-2002, the Head Start Program Information Report queried local programs about the curricula they use. In descending order of frequency, center-based programs were most likely to use the Creative Curriculum, High/Scope, a locally designed curriculum, and High Reach (beyond these a variety of other curricula were used.) This chapter examines the following: the curricula Head Start programs in the FACES sample are using; the training and ongoing support teachers receive in the use of their curricula; teacher satisfaction with their curricula; and the relationships between the type of curricula used and child, family, program, and classroom characteristics. The relationship between curricula and Head Start classroom quality is discussed in Chapter IV. Chapter V discusses the relationships between curricula and children's cognitive gains and social development in Head Start.

## RESEARCH QUESTIONS

In this chapter, the following research questions will be addressed:

1. What percentage of Head Start programs use a curriculum?
2. What types of curricula are used in Head Start programs?
3. From whom do Head Start teachers receive training and ongoing support in the use of their curriculum?
4. What percentage of Head Start teachers have access to a copy of their curriculum?
5. What aspects of the curricula do Head Start teachers like?
6. What is the relationship between the type of curricula used and the characteristics of children and families served?
7. Are there regional and rural-urban differences in the type of curricula used by Head Start programs?
8. What is the relationship between the type of curricula used and classroom quality?

## METHODS

The sample for this chapter includes 231 center-based Head Start teachers from 43 Head Start programs in the FACES 2000 study. Field staff conducted personal interviews with these teachers. The monthly family income data (N = 1,859) and child ethnicity data (N = 1,988) come from parent interviews conducted in Fall 2000. These parent interview numbers are less than the total number of parents interviewed in Fall 2000 (N =2,488) because of non-response or missing data on monthly family income or child ethnicity variables.

Weighted percentages, correlations, independent-sample t-tests and multivariate analyses of variance were used to answer the research questions. Data presented in this chapter are weighted to represent the universe of Head Start programs.

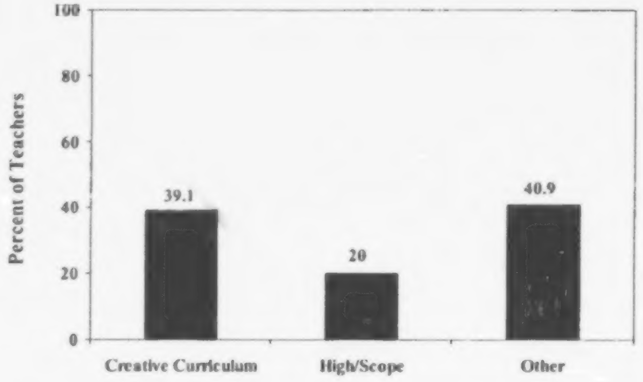
## FINDINGS

### A. Head Start Teachers Report Using a Specific Curriculum, Receiving Training, and Having Access to Their Curriculum

#### Types of Curricula

Head Start teachers were asked if they used a single specific curriculum, a combination of curricula, or no curriculum. About 70 percent of the teachers used a single curriculum, 21 percent used a combination of curricula, and 9 percent did not use a curriculum. Teachers who reported using a single curriculum or a combination of curricula were asked to name their principal curriculum. The majority (59.1 percent) said that they used either Creative Curriculum or High/Scope (see Figure 3.1). Almost 41 percent used a curriculum other than Creative Curriculum or High/Scope, which will be referred to as Other curricula. Other curricula mentioned by teachers were High Reach, Scholastic, Los Cantos Los Ninos, R.E.A.L, Global Curriculum, Creating Child Centered Classrooms – Step by Step, Building Bridges, Northern Kentucky Curriculum, Teacher Planning Wellbook, Therapeutic Intervention Program, Kid College Curriculum, Newport, and Montessori. Some said they used the "Head Start" curriculum or "theme units" (although there is no "official" Head Start curriculum).

Figure 3.1: The Majority of Teachers Used Either Creative Curriculum or High/Scope



### Curricular Training and Support

Teachers were asked if they received training in their curriculum and who provided this training (teachers could respond with only one source). Ninety-three percent reported receiving training in their curricula. Of those that received training, most had received it from their own program staff (58.5 percent), followed by curricula developers (14 percent), a Head Start Quality Improvement Center or HSQIC (10.3 percent), another Head Start program (5 percent), a University School of Education (4.4 percent), or another source (7.9 percent).

Almost 92 percent of the teachers received ongoing support in the use of their curriculum and this support could come from several sources. Of those who received support, it most often came from their supervisor or the education coordinator (70.3 percent), other teachers (23.9 percent), the HSQIC (20 percent), curriculum developers (19.3 percent), or a mentor/master teacher (14 percent). Support also came from other Head Start programs (10.8 percent), Schools of Education (6.6 percent), the disability services quality improvement center or DSQIC (4.1 percent), and other sources (8.6 percent).

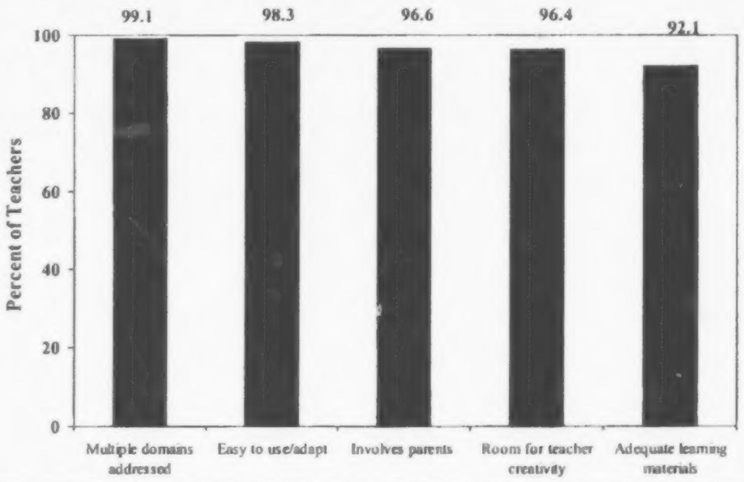
### Teacher Access to Curricula

Ninety-seven percent of the teachers responded that teachers and assistant teachers in their program had access to a copy of their curriculum.

### Teacher Satisfaction With Curricula

The majority of the teachers (92.3 percent) said that they liked their curriculum. Figure 3.2 indicates that an overwhelming majority of the teachers said they liked their curriculum because it addressed multiple domains of learning (99.1 percent), was easy to use and adapt (98.3 percent), involved parents (96.6 percent), had room for teacher creativity (96.4 percent), and had adequate learning materials/resources/examples of activities (92.1 percent).

Figure 3.2: Teachers Liked Their Respective Curricula for a Variety of Reasons



## B. Relationship Between Curricula and Family, Program, and Classroom Characteristics

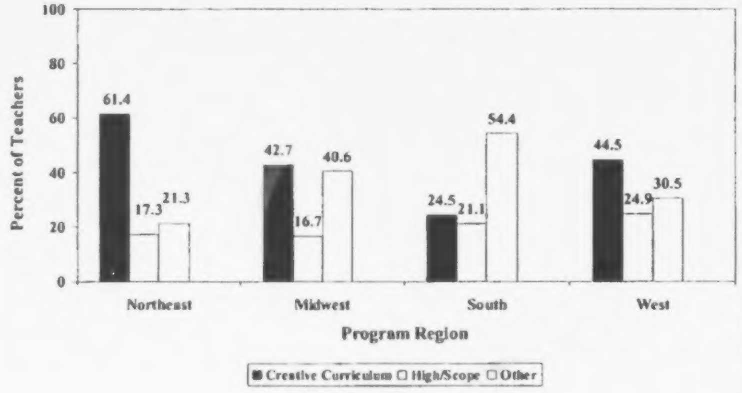
### Relationship Between Curricula and Family Characteristics

In the context of the study conceptual framework described in Chapter IV, the relationship between the use of particular curricula and the characteristics of the families served by the programs were examined. The mean monthly family income of families from classrooms using the High/Scope curriculum was $1,641.18, from classrooms using Creative Curriculum was $1,559.60, and from classrooms using Other curricula was $1,319.43. This indicates that classrooms using Other curricula served the poorest families compared to classrooms using Creative Curriculum or High/Scope. The percentage of non-white children served by teachers using Other curricula was 75.8 percent, for teachers using High/Scope it was 69.6 percent and for teachers using Creative Curriculum it was 48.8 percent. Thus, teachers using Other curricula served the highest percentage of non-white children.

### Relationship Between Curricula and Program Characteristics

Figure 3.3 shows that more teachers from the Northeast (78.7 percent), Midwest (59.4 percent), and West (69.4 percent) used Creative Curriculum or High/Scope than Other curricula while the majority of teachers from the South used Other Curricula (54.4 percent).

**Figure 3.3: Most Teachers From Northeast, Midwest, and West Used Creative Curriculum or High/Scope While Majority of Teachers From the South Used Other Curricula**



The majority of urban teachers (51.4 percent) used either Creative Curriculum or High/Scope while the majority of rural teachers (57.5 percent) used Creative Curriculum. As shown in Table 3.1, although urban and rural teachers were equally likely to use the High/Scope curriculum (20 percent), urban teachers were more likely than rural teachers to use Other curricula (48.6 percent versus 22.4 percent).

**Table 3.1: The Majority of Urban Teachers Used Either Creative Curriculum or High/Scope While the Majority of Rural Teachers Used Creative Curriculum**

| | Creative Curriculum | High/Scope | Other |
|---|---|---|---|
| Urban | 31.4 | 20 | 48.6 |
| Rural | 57.5 | 20 | 22.4 |

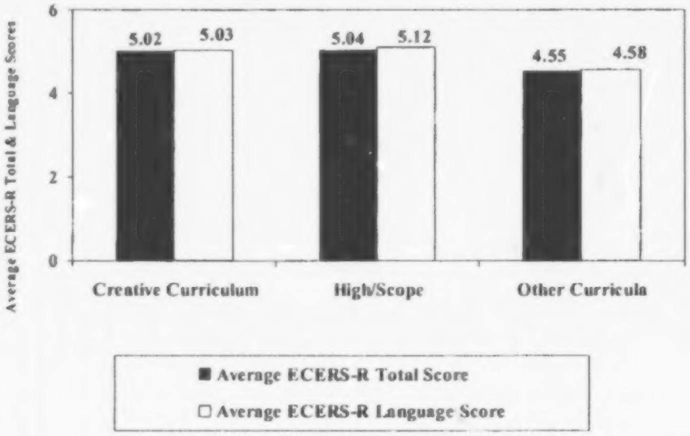
### Relationship Between Curricula and Classroom Quality

The relationship between curricula and classroom quality was examined using three different scores of classroom quality – ECERS-R Total score, ECERS-R Language Score, and Quality Composite score.

The ECERS-R Total score is derived from the Early Childhood Environment Rating Scale-Revised (ECERS-R), which provides a global rating of classroom quality based on structural features of the classroom. Scores can range from 1 (inadequate) to 7 (excellent). As displayed in Figure 3.4, the average ECERS-R Total scores for classrooms using Creative Curriculum (5.02) and High/Scope (5.04) were significantly higher ($p < .05$) than those for classrooms using Other curricula (4.55). However, the average ECERS-R Total scores for High/Scope and Creative Curriculum were not significantly different from each other.

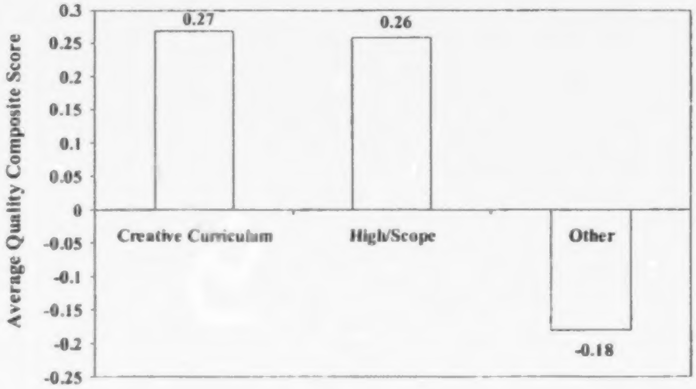
The ECERS-R Language score is a subscale of the ECERS-R and assesses classroom quality as it pertains to encouraging language-reasoning experiences (Figure 3.4). Scores can range from 1 (inadequate) to 7 (excellent). Classrooms of teachers who used Creative Curriculum (5.03) and High/Scope (5.12) had significantly higher average ECERS-R Language scores ($p < .05$) than classrooms of teachers using Other curricula (4.58). However, the average ECERS-R Language scores of classrooms that used Creative Curriculum (5.03) and High/Scope (5.12) were not significantly different from one another.

**Figure 3.4: The Classrooms of Teachers Who Used Creative Curriculum or High/Scope Had Significantly Higher Average ECERS-R Total Scores and ECERS-R Language Scores Than Classrooms of Teachers Who Used Other Curricula**



The Quality Composite score is derived from a principal components factor analysis of the ECERS-R Language score, Assessment Profile Scheduling raw score, and Assessment Profile Learning Environment raw score. The standardized factor scores were calculated, so that the scores are in standard deviation units, with a mean of 0 and a standard deviation of 1. As displayed in Figure 3.5, the classrooms of teachers who used Creative Curriculum (0.27) and High/Scope (0.26) had significantly higher Quality Composite scores ($p < .05$) than classrooms of teachers using Other curricula (-0.18). However, the Quality Composite scores of classrooms that used Creative Curriculum (0.27) and High/Scope (0.26) were not significantly different from one another.

**Figure 3.5: The Classrooms of Teachers Who Used Creative Curriculum and High/Scope Had Significantly Higher Quality Composite Scores Than Classrooms Using Other Curricula**



## Other Curricula and Classroom Quality

In order to determine the relationship of curricula within the "Other" category to classroom quality, "Other" curricula have been further categorized as Widely Available Curricula and All Other Curricula. Widely Available Curricula include High Reach, Scholastic Curriculum, Newport Curriculum and Montessori. These curricula appear to be established (for example High Reach has been commercially available for about 17 years, Scholastic for more than 82 years, and Montessori for 96 years), have information about the curricular goals and practices, printed materials and in some cases have research available on their efficacy. All Other curricula includes curricula not specified by teachers, as well as curricula termed "Head Start", Los Cantos Los Ninos, Theme Units, R.E.A.L, Global Curriculum, Creating Child Centered Classrooms – Step by Step, Building Bridges, Early Childhood Lesson Plan, Curriculum Workbook, Northern Kentucky Curriculum, Teacher Planning Wellbook, Therapeutic Intervention Program and Kid College Curriculum. This group of curricula appears to be designed by programs locally or from sources not widely available.

The classroom quality scores for Widely Available curricula, All Other curricula, Creative Curriculum and High Scope are presented in Table 3.2. The average ECERS-R Total scores and average ECERS-R Language scores for Creative Curriculum, High Scope and Widely Available Curricula are similar although the Quality Composite score for Widely Available Curricula is lower (See Figures 3.6 and 3.7).

**Table 3.2: Type of Curricula and Classroom Quality Scores (N = 228)**

| ***Type of Curricula Used*** | **Weighted Percentage of Head Start Teachers** | **Average Total ECERS-R Score** | **SD** | **Average ECERS-R Language Score** | **SD** | **Quality Composite Score** | **SD** |
|---|---|---|---|---|---|---|---|
| Creative Curriculum | 39.1 | 5.02 | .92 | 5.03 | 1.32 | 0.27 | .96 |
| High/Scope | 20.0 | 5.04 | .73 | 5.12 | 1.01 | 0.26 | .84 |
| Widely Available Curricula | 9.8 | 4.82 | .63 | 5.15 | .96 | 0.07 | .95 |
| All Other Curricula | 31.1 | 4.47 | .86 | 4.4 | 1.19 | -0.27 | .95 |

- Widely Available Curricula includes: High Reach, Scholastic Curriculum, Newport Curriculum & Montessori.
- All Other Curricula includes: curriculum not specified, Head Start curriculum, Los Cantos Los Ninos, Theme Units, R.E.A.L, Global Curriculum, Creating Child Centered Classrooms – Step by Step, Building Bridges, Early Childhood Lesson Plan, Curriculum Workbook, Northern Kentucky Curriculum, Teacher Planning Wellbook, Therapeutic Intervention Program, and Kid College Curriculum

**Figure 3.6: Classrooms That Used Creative Curriculum and High/Scope Had Significantly Higher Average ECERS-R Total Scores and ECERS-R Language Scores Than Classrooms That Used All Other Curricula.**

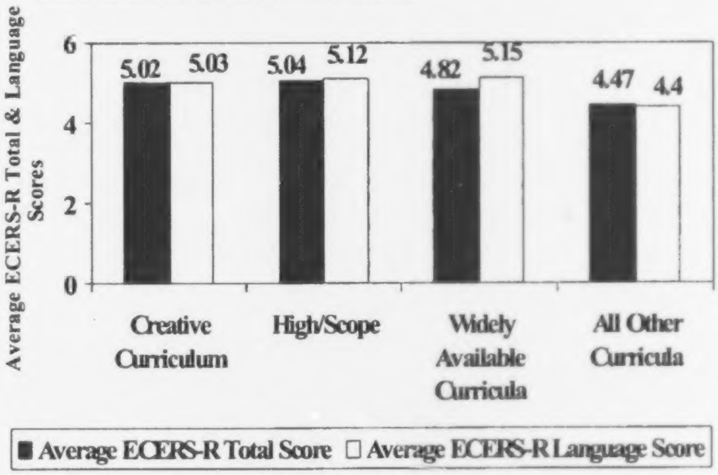
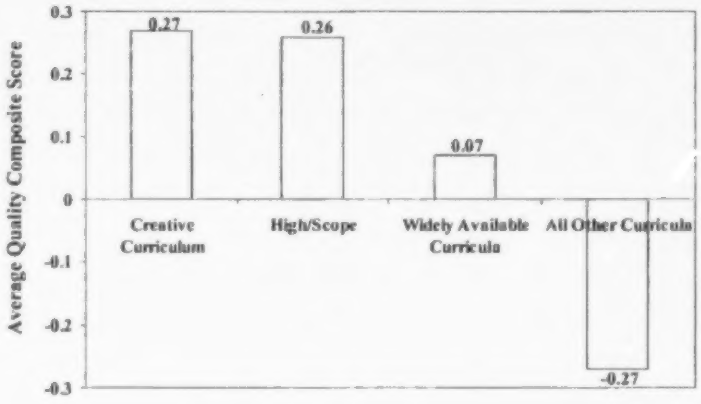


**Figure 3.7: Classrooms That Used Creative Curriculum and High/Scope Had Significantly Higher Quality Factor Scores Than Classrooms That Used All Other Curricula.**



Multivariate analyses of variance were used to examine if the classroom quality scores were significantly different between classrooms using Creative Curriculum versus High/Scope, Creative Curriculum versus Widely Available curricula, Creative Curriculum versus All Other curricula, High/Scope versus Widely Available curricula, High/Scope versus All Other curricula, and Widely Available Curricula versus All Other curricula. Significant differences (p < .05) were found between classrooms using Creative Curriculum versus All Other on all three classroom quality scores; and between classrooms using High/Scope versus All Other on all three classroom quality scores. Classroom quality scores between classrooms using Widely Available Curricula versus All Other curricula were not significantly different.

- Classrooms using the Creative Curriculum when compared to classrooms using All Other curricula had significantly higher average ECERS-R Total scores (5.02 versus 4.47), average ECERS-R Language scores (5.03 versus 4.4), and Quality Composite scores (0.27 versus - 0.27).
- Classrooms using the High/Scope curriculum when compared to classrooms using All Other curricula had significantly higher average ECERS-R Total scores (5.04 versus 4.47), average ECERS-R Language scores (5.12 versus 4.4), and Quality Composite Scores (0.26 versus - 0.27).
- Classrooms using Widely Available Curricula when compared to classrooms using All Other curricula were not significantly different. Classrooms using Widely Available Curricula were also not significantly different from those using Creative Curriculum, or those using High/Scope. Classrooms using Creative Curriculum were not significantly different from those using High/Scope.

Overall, these findings suggest that classrooms using curricula such as Creative Curriculum and High/Scope have relatively higher classroom quality than Head Start classrooms using other curricula.

## IMPLICATIONS

- The Head Start Performance Standards mandate that programs use a curriculum although curricula are not prescribed. Findings indicate that the great majority of the Head Start programs comply with this standard consistent with the goal of providing a planned, developmentally appropriate early childhood program for children.
- There is a relationship between program characteristics (region, urban-rural, characteristics of children/families) and whether or not a program uses a curriculum and the type of curriculum used.
- The relationship between curricula and classroom quality may reflect the influences of other factors (such as the resources available to programs for purchasing and training in specific curricula), or may demonstrate the effect of certain curricula on quality (see Chapter IV). Programs may want to consider whether using curricula related to higher levels of classroom quality would enhance their programs.

# Chapter IV: Understanding Quality in Head Start Classrooms

Reports on the first cohort of FACES have demonstrated that Head Start is providing children with high-quality educational services as measured by standard observational instruments. In this chapter, we present results of quality analyses from the second FACES cohort (FACES 2000), consisting of a new national sample of 43 Head Start programs and classrooms, collected during the fall of 2000 and the spring of 2001. This chapter describes the nature of quality in Head Start programs and the factors that help explain variations in quality across Head Start classrooms. Findings about the relationship between classroom quality and child outcomes are described in Chapter V.

## RESEARCH QUESTIONS

Information collected through classroom observations and interviews with Head Start teachers and parents were analyzed to answer the following research questions:

1. What is the quality of Head Start classrooms in 2000-2001, and how does it compare to quality reported in 1997-1998?
2. What are the backgrounds, qualifications and experiences of Head Start teachers in 2000-2001 and were there any changes from the earlier cohort (1997-1998)?
3. What is the relationship between the quality of Head Start classrooms and teacher qualifications, experience, and attitudes and knowledge; which of these factors seems most important in explaining variations in quality?
4. Do factors beyond the classroom, such as the types of curricula Head Start programs provide, the average teacher salaries, and characteristics of families served by the Head Start program explain variations in the quality of Head Start classrooms?

### A. What Quality Means and How We Appraise It

A variety of indices were used to measure quality in FACES because of the different elements of quality cited in the research literature (Phillips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000). In addition to "structural" indices, such as staff-child ratio (Cost, Quality and Outcome Studies Team, 1995; Howes, Phillips, & Whitebook, 1992) and group size (Ruopp, Travers, Glantz, & Coelen, 1979), the backgrounds and experiences of early childhood teachers are important (Cost, Quality and Outcome Studies Team, 1995). Additionally, a variety of "process" aspects of the classroom environment, including teacher-child interactions, child-child interactions, as well as curriculum, the schedule of activities and materials have been considered important (Bredekamp, 1986; Hayes, Palmer, & Zaslow, 1990; Phillips & Howes, 1987; Scarr & Eisenberg, 1993). Recent research has also shown that factors "beyond the classroom door," which are features of the centers in which classrooms are located, may have potentially powerful influences on children's experiences of quality care in early childhood classrooms (Blau, 1997; Hofferth & Chaplin, 1998 cited in Phillips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000).

In FACES, quality was considered to include not only the number of children and adults in each classroom, but process factors such as the availability of learning materials, the types of classroom activities, the scheduling, and the variety of learning opportunities provided to all children. We also interviewed lead teachers[11] in Head Start classrooms to collect teacher background information (experience and qualifications) as well as more detailed information about their curriculum, classroom activities, and attitudes and knowledge about early childhood education practices.

The classroom quality measures used in FACES 2000 provide direct evidence for the extent to which Head Start programs employ qualified teachers and provide rich learning environments and curricula for the children. In FACES 2000, we used the same measures as those employed in the earlier cohort, in order to make some comparisons in quality, and we added several key measures. The classroom quality measures are fully described in the Appendix, and will only be briefly mentioned here.

*The Early Childhood Environment Rating Scale (ECERS-R):* This revised version (Harms, Clifford, & Cryer, 1998) consists of 37 scales measuring a wide variety of quality related processes occurring in the classroom, including: routines; teacher-child interaction, particularly in the use of language; learning activities; classroom tone, creative, dramatic, and gross and fine motor activities; equipment and furnishings; and staff and parent facilities. A high score on the total ECERS-R indicates higher classroom quality, in terms of equipment, space and play materials, as well as the range of activities and staff-child interactions.

*The ECERS-R Language Subscale:* This measure consists of four items from the ECERS-R that assess the quality of the language environment in Head Start classrooms. A high score indicates a classroom with a rich language environment, in terms of the availability and use of books and printed materials, receptive and expressive language activities, language to engage logical and reasoning skills, and the informal use of language throughout the classroom day.

*The Assessment Profile for Early Childhood Programs: Research Edition Scheduling scale.* This scale from the Assessment Profile for Early Childhood Programs (Abbott-Shim & Sibley, 1998) assesses the written plans for classroom scheduling and how classroom activities are implemented. A high score indicates that the teacher takes a planful approach to the classroom schedule and has been intentional regarding a variety of socialization and learning experiences for children throughout the day.

*The Assessment Profile for Early Childhood Programs: Research Edition Learning Environment scale.* This scale from the Assessment Profile for Early Childhood Programs (Abbott-Shim & Sibley, 1998) measures the variety of learning materials available and accessible in the classroom that provide learning experiences in different developmental areas. It also assesses the degree to which the classroom provides for a "language-rich" environment through language learning materials as well as the labeling of objects, and the amount of printed material in the classroom. A high score on this scale is indicative of a greater variety of materials accessible, that stimulate growth in all developmental domains.

*The Assessment Profile for Early Childhood Programs: Research Edition Individualizing scale.* This is based on a scale from the Assessment Profile for Early Childhood Programs

---

[11] The results of teacher data presented in this chapter were based on interviews and ratings of the lead or senior teacher in each classroom.

(Abbott-Shim & Sibley, 1998). For FACES 2000 it has been revised and shortened to five observational items measuring how the teacher plans the classroom activities to meet the varying learning needs of each child, how the teacher keeps track of the children's work during the year through the use of individual child portfolios and how the teacher is able to accommodate children with disabilities. A high score indicates that teachers are able to adjust classroom activities to meet the learning needs of individual children.

*The Arnett Caregiver Interaction Scale (Arnett, 1989).* This rating scale consists of 26 items that measure the teacher's sensitivity, punitiveness, detachment, permissiveness, and encouragement of child independence and self-help skills. A high score indicates greater teacher sensitivity, responsiveness and encouragement of children's independence and self-help skills, and lower levels of punitiveness and detachment.

*Teacher Interview.* The teacher from each classroom was asked specific questions about the nature of the curriculum used, attitudes and knowledge about early childhood education practice, how they monitor the progress of individual children, and what accommodations the teacher makes to meet the learning needs of each student, including those with special needs. The interview also collected extensive information about the teachers' backgrounds (e.g., age, ethnicity), experience (e.g., total years teaching, years teaching Head Start), and qualifications (e.g., whether the teacher has a BA or AA, whether the teacher had some graduate school education, whether the teacher has a Child Development Associate, the course of study, and licensure). Ethnicity was included in these analyses because it may be linked to differences in teacher qualifications and experience and because the types of teachers in the classrooms may be influenced by the backgrounds of the families and children attending the Head Start program as well as the larger community served by the program.[12]

*Teacher Attitudes and Knowledge.* The 24-item Teacher Beliefs Scale (Burts, Hart, Charlesworth, & Kirk, 1990) was included in the teacher interview, and consists of statements worded to reflect positive attitudes and knowledge of generally accepted practices in preschool settings, or to reflect a lack of these attitudes and knowledge. In FACES 2000 we used one factor comprising 9 items that explained most of the variation in scores for the entire scale. A high score indicates higher positive attitudes and knowledge about early childhood education practices.

*Quality Composite Score.* We found that several of the key quality indicators were highly correlated with each other, suggesting that for analytical purposes we can explain a greater amount of variation in quality by reducing the three indicators in question to one measure. Scores from the ECERS-R Language Scale score and the Assessment Profile Scheduling and Learning Environment were combined to form this single score for quality. A higher score indicates higher levels of quality.

*Child-Adult Ratio.* Classroom observers counted the number of children, the number of adults, and the number of paid staff at two time periods during the classroom day. The two occasions were separated by at least one hour and involved one structured (teacher-directed) and one unstructured activity. A higher child-adult ratio is indicative of lower quality.

---

[12] While age was also included in the teacher interview, it was so highly correlated with the teacher's years of experience that it was not included in analyses beyond the descriptive level. However, since ethnicity did prove to be related to other factors in classroom quality, we continued to include it in our analyses.

During the Fall 2000 data collection period, which operated from September 2000 to December 2000, we collected classroom quality and teacher data from 225 Head Start centers in 43 programs around the country. We observed 278 classrooms out of 286 possible for a completion rate of 97 percent.[13] Percent agreement between two independent observers in a sample of classrooms averaged 93.5 percent for the Assessment Profile Scheduling Scale, 87.9 percent for the Assessment Profile Learning Environment Scale, and 86.7 percent for the Assessment Profile Individualizing Scale. Percent agreement across all ECERS scales (which includes direct hits and being off by one on a seven-point scale) averaged 79.5 percent, agreement on the ECERS-R Language subscale averaged 85.7 percent, and agreement on the Caregiver Interaction Scale averaged 93.9 percent.

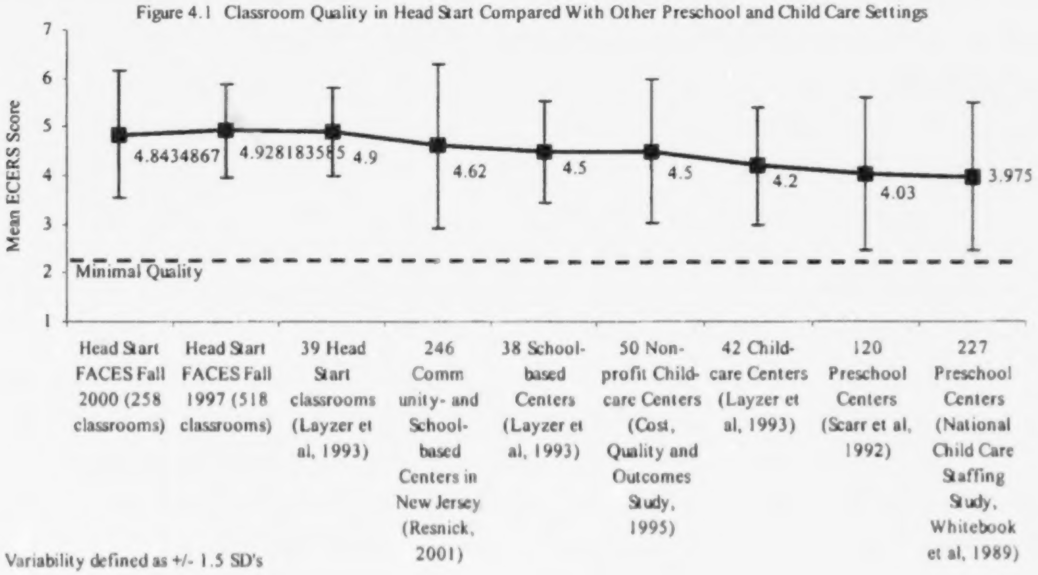
---

[13] Due to some missing data among one or more of the quality indicators, the analytic sample was comprised of 258 classrooms.

## B. The Quality of Most Head Start Classrooms Continues To Be Good

In the first FACES cohort, Fall 1997 to Spring 1998, we reported that the quality of Head Start classrooms was good and above that usually found among center-based preschools. In FACES 2000, using a new sample of 43 programs, we continue to find that quality is good and above that found among other center-based preschools. Figure 4.1 compares the two FACES cohorts with other studies using the overall ECERS score as the measure of quality. Not only is the average score higher, but the variation is less than that found in many other preschool settings.14

Figure 4.1 Classroom Quality in Head Start Compared With Other Preschool and Child Care Settings



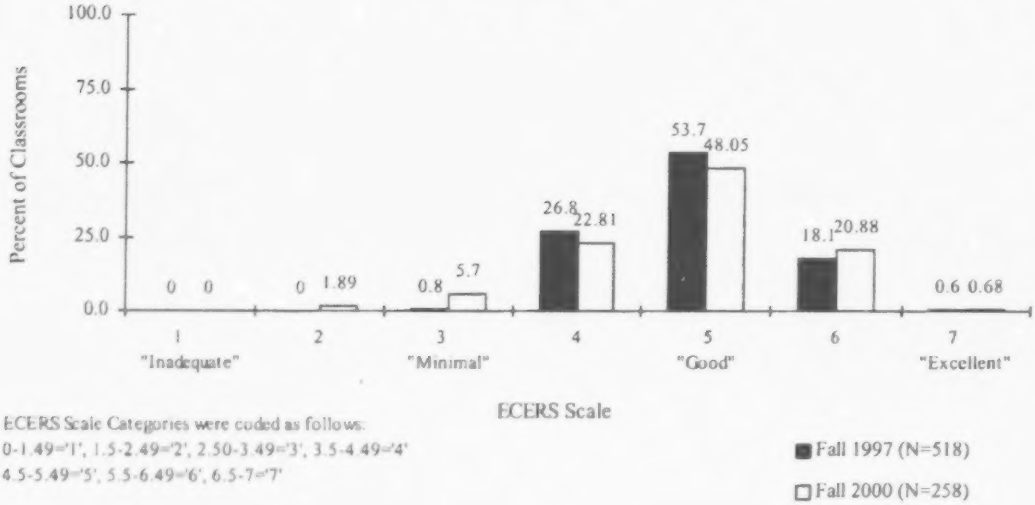
Variability defined as +/- 1.5 SD's

**The Early Childhood Environment Rating Scale (ECERS).** In the new cohort of 43 programs, the overall average ECERS-R score for the 258 classrooms observed in Fall 2000 was 4.84 (with a standard deviation of .87). This compares with an overall average ECERS score for the 518 classrooms in the first FACES cohort (Fall 1997) of 4.93 (with a standard deviation of 0.63). While the FACES 2000 average score is comparable to that found earlier, the variability in the scores was higher, and approached that reported by the test developers for the revised measure. Overall, the average scores are consistent and indicate that Head Start classrooms continue to show good quality.

---

[14] It should be noted that, with the exception of the New Jersey study, all other studies used the original ECERS, and at this writing there are no published reports attesting to the appropriateness of comparing scores on the two versions.

Figure 4.2 Distribution of Classrooms on ECERS and ECERS-R, Fall 1997 and 2000



ECERS Scale Categories were coded as follows:
0-1.49='1', 1.5-2.49='2', 2.50-3.49='3', 3.5-4.49='4'
4.5-5.49='5', 5.5-6.49='6', 6.5-7='7'

Figure 4.2 compares the FACES 1997 and FACES 2000 cohorts on the ECERS total score, with 1 indicating "inadequate" quality and 7 indicating "excellent" quality. In FACES 1997, no classrooms were scored in the "inadequate" range (1 or 2) and only 4 of 518 classrooms were scored in the "minimal" range (score of 3). In FACES 2000, using the revised version of the ECERS which more strictly differentiated the highest and lowest ends of the scale, 5 of 258 classrooms (1.9 percent) were scored in the "inadequate" range and 15 of 258 classrooms (5.7 percent) scored in the "minimal" range. Thus, compared to the earlier cohort, in FACES 2000 a small number of classrooms were rated lower in quality, but low-scoring classrooms still represented only 20 of 258 classrooms overall (7.6 percent). On the other hand, the number of classrooms rated "excellent" (ECERS scores of 6 or higher) increased from FACES 1997 to FACES 2000. In FACES 1997, there were 97 out of 518 (18.7%) classrooms rated "excellent" while in FACES 2000, 21.6 percent of classrooms (56 out of 258) were rated "excellent" (scores of 6 or higher).

**Assessment Profile Scheduling, Learning Environment, and Individualizing Scales.** On the Scheduling scale, the raw scores were virtually identical from FACES 1997 to FACES 2000 (average raw scores of 11.17 in 1997 and 11.12 in 2000). On the Learning Environment scale, there was a slight but non-statistically significant increase in raw scores over the two cohorts, from 13.46 in 1997 to 14.44 in 2000. Finally, on the Individualizing Scale, a new measure in FACES 2000, the average raw score was 3.58 (out of a maximum raw score of 5), with a standard deviation of 1.2, indicating that overall Head Start classrooms provide an environment that takes into account the learning needs of individual students, but there is room for improvement. In particular, on the Individualizing Scale, 60 percent of classrooms maintained portfolios on individual children but only 37 percent of classrooms provided opportunities for children to evaluate their work, or to decide which products are included in their portfolios. However, with regard to the inclusion of children with disabilities, and making accommodations to allow these children to be included in classroom activities, 94 percent of classrooms were rated as having full inclusion and 90 percent of classrooms had provisions for accommodating children with special needs.

**Arnett Caregiver Interaction Scale.** On this scale, measuring the sensitivity and responsiveness of teachers in Head Start classrooms, the average score in FACES 2000 was 71.5, almost identical to the score of 71.3 in FACES 1997.

The fact that the scores for three measures of classroom quality, the ECERS, the Assessment Profile measures and the Arnett Caregiver Interaction Scale, were consistent across both cohorts adds support to the conclusion that, in general, quality in Head Start classrooms was consistent over the two cohorts of classrooms, from 1997 to 2000.

**Child-Adult Ratio.** The average child-adult ratio for the FACES Head Start classrooms in Fall 2000 was 5.4 children per adult, compared with 6.3 children per adult at the Fall 1997 observation and 6.2 children per adult during the Spring 1998 observation.[15] Looking only at paid staff, Fall 2000 classrooms averaged 6.5 children per paid staff. When we include other adults assisting paid staff in classroom activities, the child-adult ratio averaged 5.4, which is somewhat lower than that found in Fall 1997. The difference in these ratios suggests the important influence of volunteers on improving child-adult ratios in Head Start classrooms (lower ratios indicate higher quality). These ratios are far better than the NAEYC accreditation standard of eight or fewer three-year-olds or 10 or fewer four-year-olds for each adult and exceed the Head Start Program Performance Standards of 7.5 to 8.5 or fewer three-year-olds or 10 or fewer four-year-olds per adult.

**Teacher Backgrounds, Qualifications, and Experience.** In Fall 2000, we continue to find that Head Start teachers overall are experienced and qualified (see Table 4.1). Teachers in Head Start classrooms have been teaching in Head Start for 7.9 years and they have been teaching for an average of 11.8 years in all educational settings. These data are almost identical to those reported for the earlier cohort, Fall 1997. Teachers spent most of their teaching careers in Head Start classrooms but there was a wide range of teaching experience. Approximately 21 percent of the Head Start teachers were relatively new, having been teaching in Head Start for less than two years, and 28 percent had taught in Head Start for 10 years or more.

Compared to the Fall 1997 cohort, there were several statistically significant differences.[16] More new teachers (21 percent) taught in Fall 2000, compared with Fall 1997 when only 14 percent had taught Head Start for less than 2 years. Also, the percentage of teachers who taught from five to nine years declined from 34 percent in Fall 1997 to 28 percent in Fall 2000. Approximately the same number of teachers in both cohorts had been teaching in Head Start for ten years or more (28 percent in Fall 2000 and 29 percent in Fall 1997).

---

[15] In Fall 1997 we only measured total number of adults in the classroom so that the discrepancy between 1997 and 2000 could be due to the slightly different measurement methods. In the remaining analyses in this chapter, we use the child-adult ratio rather than the child-staff ratio, to remain consistent with the earlier cohort. In fact, the two measures are so highly correlated that results of statistical analyses using each will be almost identical.

[16] The chi-square comparison using unweighted data on the independent samples was statistically significant at $p<.05$. Unweighted data were used because no classroom weights were created for the fall 1997 sample.

Most Head Start teachers have good teaching qualifications, but lower than those of teachers in public elementary schools. In a survey of pre-kindergarten classrooms in the U. S. public schools in 2000-2001, 86 percent of pre-kindergarten teachers had a bachelor's or higher degree (Smith, Kleiner, Parsad, & Farris, 2002). In the FACES Fall 2000 cohort, 27.8 percent had a bachelor's degree, 18.6 percent had an associate's degree, and another 32.2 percent had some college but no degree. Overall, 46.4 percent of teachers had either a bachelor's or an associate's degree with 38.7 percent having a bachelor's degree or higher and 57 percent having an associate's degree or higher. Seventy-four percent of all teachers reported having the Child Development Associate (CDA) credential or a state-awarded preschool certificate, with 58 percent having the Child Development Associate only.

The proportion of teachers with a bachelor's degree or higher increased significantly from 28.1 percent in Fall 1997 to 38.7 percent in Fall 2000, primarily due to an increase in the proportion of teachers with graduate level degrees, defined as a Master's degree, its equivalent or higher. In Fall 2000, 10.9 percent of teachers reported having these advanced degrees compared with only 3.2 percent in Fall 1997, and in Fall 2000 sixteen percent of teachers reported having received some graduate level training.[17] This increase was significantly correlated with the increase in the proportion of new teachers in Head Start, from 14 percent to 21 percent over the same time period. In Fall 2000, 32.9 percent of teachers with graduate level education or higher were new teachers, compared with only 13.3 percent in Fall 1997.[18] These results suggest that there is an increased number of new teachers with advanced degrees entering Head Start.

---

[17] The chi-square comparison using unweighted data on the independent samples was statistically significant at $p<.01$.

[18] In Fall 2000 the chi-square test using weighted data was statistically significant at $p<.001$. Further, a comparison of the two proportions from Fall 1997 to Fall 2000 using unweighted data also revealed a statistically significant increase at $p<.01$ ($z = 4.441$).

Table 4.1 Comparison of Lead Teacher Backgrounds, FACES Fall 1997 and Fall 2000

| | | **Fall 1997** (N=437) | **Fall 2000** (N=257) |
|---|---|---|---|
| Years Teaching Head Start | 1-2 YRS | 14.2% | 21.1% |
| | 3-4 YRS | 22.7% | 23.5% |
| | 5-9 YRS | 34.1% | 27.6% |
| | 10+ YRS | 29.0% | 27.8% |
| | Total | 100% | 100% |
| Highest Level of Education Achieved | High School or Equivalent | 10.8% | 10.4% |
| | Some College | 31.4% | 32.2% |
| | Associates Diploma | 29.7% | 18.6% |
| | Bachelor's Degree or equivalent | 24.9% | 27.8% |
| | Graduate or Professional Degree | 3.2% | 10.9% |
| | Total | 100% | 100% |
| Teacher Age Category | 18-29 | 14.7% | 14.9% |
| | 30-39 | 33.3% | 33.4% |
| | 40-49 | 31.8% | 28.3% |
| | 50-59 | 15.9% | 16.3% |
| | 60 or Older | 4.3% | 7.1% |
| | Total | 100% | 100% |
| Membership in Early Childhood Education Association | No | 47.1% | 38.0% |
| | Yes | 52.9% | 62.0% |
| | Total | 100% | 100% |
| CDA CERTIFICATE/CREDENTIAL? | No | 23.9% | 26.0% |
| | Yes | 76.1% | 74.0% |
| | Total | 100% | 100% |
| Teacher Ethnicity | Caucasian, non-Hispanic | 41.1% | 48.1% |
| | African-American, non-Hispanic | 34.2% | 33.7% |
| | Hispanic* | 22.4% | 15.4% |
| | Asian | 2.3% | 1.2% |
| | Multiple Race/Other | na | 1.7% |
| | Total | 100% | 52% |

* Puerto Rico was represented in FACES 1997, but not in the FACES 2000 sample, explaining the lower percentage of Hispanic teachers.

In addition to an increase in teachers with advanced degrees, more teachers are reporting having studied Early Childhood Education or Child Development in their studies for their highest degree, whether for an associate's or bachelor's degree or higher. In Fall 2000, 78 percent indicated that their field of study included Early Childhood Education or Child

Development, compared with approximately 62 percent in Fall 1997. However, the proportion of teachers having a Child Development Associate or state preschool certificate has stayed approximately the same (74 percent in Fall 2000 compared with 76 percent in Fall 1997).

There was also an increase in the membership of teachers in a national professional association for early childhood educators (e.g., NAEYC, NHSA, and NEA), from 53 percent in Fall 1997 to 62 percent in Fall 2000. Head Start teachers were, on average, 41.4 years old with a range from 23 to 73 years of age (standard deviation of 11.1 years). One third of teachers belonged to the 30 to 39 year age group and another 28 percent were between the ages of 40 and 49 years. There were no differences in the proportions of teachers in each of the ten-year age groupings between the earlier cohort (Fall 1997) and this cohort (Fall 2000), with the exception that this cohort reflects an aging of teachers at the upper end of the distribution. Whereas only 4 percent of teachers were 60 years of age or older in Fall 1997, in this newest cohort (Fall 2000), 7 percent of teachers were 60 years of age or older.

In terms of racial and ethnic background, 33.7 percent of the teachers were African American, 15 percent were Hispanic, 1 percent were Asian and 48 percent were Caucasian. Compared with the Fall 1997 data, teachers in Fall 2000 were slightly less likely to be black (the percentage of African American teachers declined to 33.7 percent from 34.2 percent in Fall 1997). Although it may appear that there was a strong decline in the percentage of Hispanic teachers, this decline is due entirely to the fact that Puerto Rico was included in the Fall 1997 but excluded from the Fall 2000 cohort. In FACES 2000, there was an increase in the proportion of teachers who were Caucasian, from 41 percent in Fall 1997 to 48 percent in Fall 2000.

In general, the data reveal that Head Start teachers are experienced and qualified to teach early childhood education. Compared with the first FACES cohort, Head Start teachers in FACES 2000 are more qualified than were teachers in 1997. They are more likely to be Caucasian, younger, new to teaching Head Start and entering with higher educational levels including graduate degrees. They are also more likely to be trained in early childhood education and to be members of a professional organization.

**Teacher attitudes and knowledge.** In Fall 2000, on average, Head Start teachers showed generally positive attitudes and knowledge about instructional practice in early childhood education, with a mean score for all teachers of 7.9 out of a maximum score of 10. For example, they tended to agree with statements indicating positive attitudes and knowledge about early childhood education practices such as: "Head Start classroom activities should be responsive to individual differences in development" and "Children should be allowed to select many of their own activities from a variety of learning areas that the teacher has prepared (writing, science center, etc.)." Head Start teachers tended to disagree with statements indicating negative attitudes and a lack of knowledge about early childhood education practices such as: "Each curriculum area should be taught as a separate subject at separate times" and "Students should work silently and alone on seatwork."

## C. Teacher Qualifications and Experience Are Related to Classroom Quality

When correlating teacher backgrounds, qualifications and experience with quality we found that teachers with more experience and higher levels of education tended to be in classrooms rated higher in classroom quality. Specifically, we found that:

- teachers with higher levels of education tend to be in classrooms rated higher on a number of quality indicators, including the ECERS-R Language subscale, the Caregiver Interaction Scale, and the ECERS-R total score;
- teachers with a BA or AA degree were in classrooms with lower child-adult ratios (indicating higher quality) but teachers with a Child Development Associate tended to be in classrooms with higher ratios, indicative of lower quality;
- teachers who had more years teaching overall (not just in Head Start) were rated higher in their sensitivity and responsiveness, as measured by the Caregiver Interaction Scale and had higher overall ECERS-R scores;
- teachers with more years teaching preschool or Head Start were in classrooms rated higher on the Quality Composite score;
- teachers who were members of an early childhood education professional association were rated higher in their sensitivity and responsiveness, as measured by the Caregiver Interaction Scale, and were in classrooms rated higher in ECERS-R Language, the Assessment Profile Individualizing Scale, with higher Quality Composite scores;
- teachers with a Child Development Associate certificate or a state-sponsored equivalent (versus no certificate) were in classrooms rated higher on the Assessment Profile Individualizing scale;
- teachers with at least some graduate school education (versus no graduate school education) were also in classrooms rated higher in quality as measured by the ECERS-R total score; and
- teachers who had a teaching certificate (versus no certificate) were in classrooms rated higher on the overall ECERS-R and the ECERS Language scale, with lower child-adult ratios (indicating higher quality).

However, years of experience teaching Head Start (rather than overall years of teaching at any level) and the teachers' total annual salaries were not significantly correlated with any of the quality indicators.

These results indicate that teacher backgrounds and qualifications are related to higher levels of quality in Head Start classrooms. However, the above results are based only on simple correlations, and do not take into account the complex interplay between the various teacher-related factors. To understand these links between factors, we must also consider how teacher backgrounds and experience are related to their attitudes and knowledge about early childhood education practice.

## D. Teacher Backgrounds and Experience Are Related to Their Attitudes and Knowledge

Teachers who had higher levels of knowledge about early childhood education practices tended to be more experienced and better educated. Specifically, teachers with higher scores for attitudes and knowledge were more likely to:

- have higher levels of educational attainment,
- have some graduate school education or higher,
- have more total years teaching,
- belong to an early childhood education association,
- teach in another language (not including Spanish or English),[19] and
- be Caucasian, Hispanic, or Asian.

Teacher's attitudes and knowledge were not significantly correlated with: teacher salary, years teaching Head Start, having a teaching certificate, having a Child Development Associate certificate, or having a course of study in early childhood education.

This relationship between teacher education and attitudes and knowledge about early childhood education practices is supported by findings from other studies. Abbott-Shim, Lambert, and McCarty (2000) also reported that teachers with higher levels of education also showed more positive attitudes and knowledge about early childhood education practices.

Our results suggest that teachers who are better trained, with higher levels of educational attainment especially at the graduate school level, with more years of teaching experience overall, and who enroll in a related professional association are more likely to have knowledge and positive attitudes about early childhood education practices. These attitudes and knowledge should be expected to influence classroom quality, particularly since teacher credentials were related to classroom quality. In the next step toward building a model explaining the connection between teacher backgrounds, qualifications and experience and classroom quality, we test the relationship between teacher attitudes and knowledge and classroom quality.

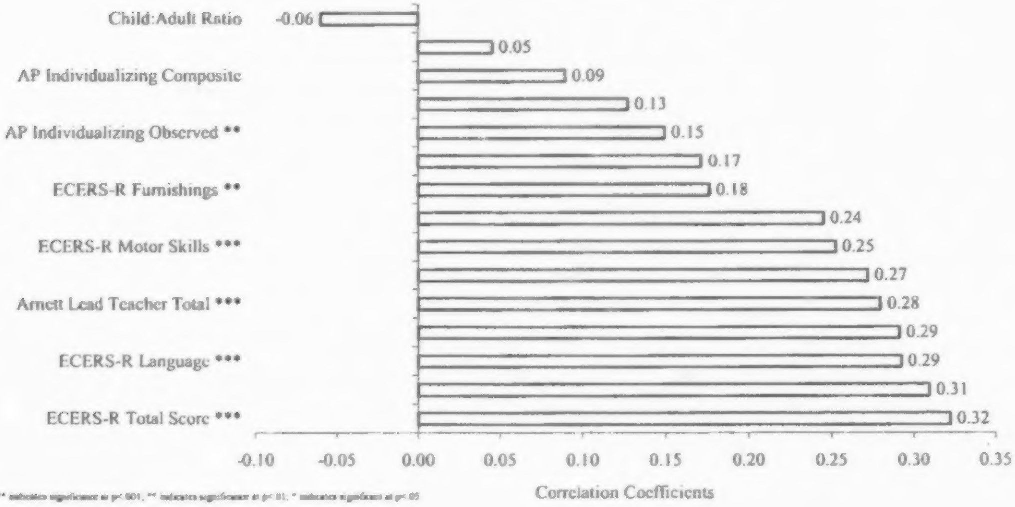
### E. Teacher Attitudes and Knowledge Are Related to Classroom Quality in Head Start

Teachers with more positive attitudes and knowledge about early childhood education practices tend to be in classrooms rated higher in quality. This relationship was found for a wide variety of quality indicators, including the Assessment Profile Learning Environment scale, the ECERS-R total score, the ECERS-R Language scale, the Assessment Profile Individualizing scale, the Quality Composite score, and the Arnett Caregiver Interaction Scale (see Figure 4.3).

---

[19] In FACES 2000 only 1 percent of teachers indicated they taught at least part of the instructional day in a language other than English or Spanish.

Figure 4.3 Teacher Beliefs Are Correlated With Classroom Quality, Fall 2000



*** indicates significance at p<.001; ** indicates significance at p<.01; * indicates significant at p<.05

Teachers holding more positive attitudes and knowledge of early childhood education practices were more sensitive and responsive to children, as measured by the Arnett Caregiver Interaction Scale. They were also more likely to adjust activities to meet the varying needs of individual children, as indicated by the Assessment Profile Individualizing scale. Teachers holding more positive attitudes and knowledge of practices in early childhood education were also in classrooms rated higher in overall quality on the ECERS-R and in classrooms with higher quality language activities, as rated by the ECERS-R Language scale.

The above findings show that teacher backgrounds, experience and qualifications, notably their level of education, are significantly related to both teacher's attitudes and knowledge about early childhood education practice, and classroom quality. The teacher attitudes and knowledge score alone was also significantly correlated with classroom quality. Thus, in predicting high or low quality in Head Start classrooms, both teacher qualifications as well as their attitudes and knowledge about early childhood education practices are important.

These results are limited because we cannot specify from the simple correlations above whether the relationships with classroom quality are independent, suggesting that teacher education, for example, and teacher attitudes and knowledge contribute separately and uniquely to variations in quality, or whether they are linked in some way. Additionally, the research literature and earlier analyses of FACES data indicate that there are factors existing at the level of the program that may influence classroom quality. The Second Performance Measures Report described how classroom quality varies across classrooms and across programs suggesting that factors at the level of the Head Start program may help to explain variations in quality. Thus, in developing a comprehensive model for understanding the many sources of influence on classroom quality, it is important to look at factors at both the classroom and the program levels.

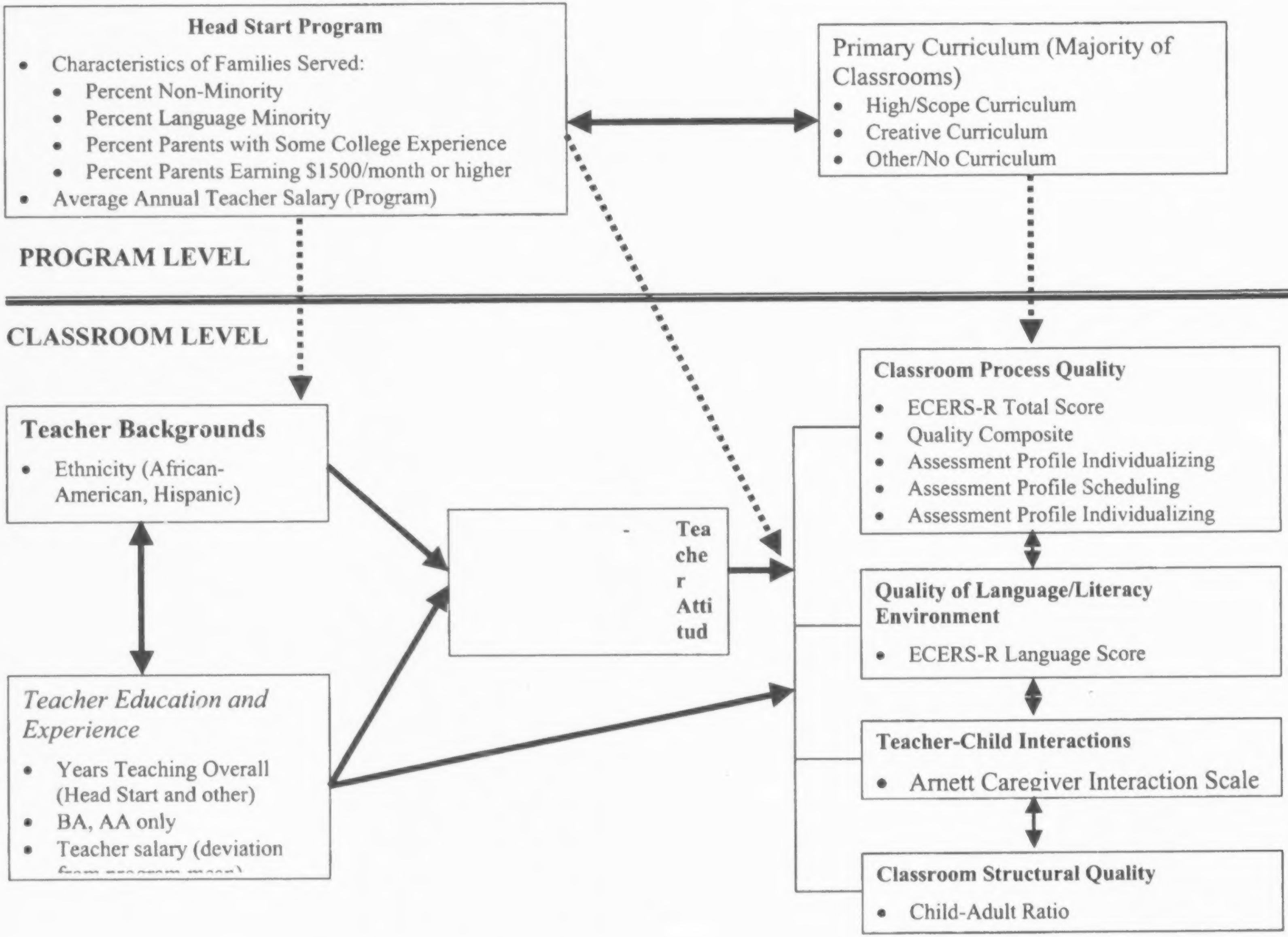
## F. Teacher and Program-Level Factors Explain Significant Variations in Classroom Quality

To determine the joint influence of both program- and classroom-level factors in predicting classroom quality, we used a multi-level approach with two levels of factors: program and classroom.[20] This approach tests a model for explaining quality in Head Start that builds from the previous analyses, which were done only at the level of the classroom, and then adds factors at the level of the Head Start program. Head Start programs consist of the grantee and delegate agencies that administer the centers and classrooms within its purview. A program comprises the administrative entity primarily responsible for determining budgets, staffing, and the allocation of resources across centers and classrooms, as well as for choosing curricula for its classrooms and providing training and resources to support the curricula. Programs also define the geographic areas they serve and identify the needs of low-income families and children living there. Thus, in order to understand variations in classroom quality we must include program-level factors. The model we are testing will explain how variations in the quality of Head Start classrooms may be due to those teacher factors identified earlier, as well as factors that exist at the level of the program, such as the characteristics of families participating in the program, the curriculum used in the majority of classrooms, and the average annual teacher salary (see Figure 4.4).

---

[20] We used the PROC MIXED statistical procedure in SAS, as outlined by Singer (1998).

Figure 4.4 A Model of Program- and Classroom-Level Factors Explaining Quality in Head Start


Head Start Program
Characteristics of Families Served:
Percent Non-Minority
Percent Language Minority
Percent Parents with Some College Experience
Percent Parents Earning $1500/month or higher
Average Annual Teacher Salary (Program)
Primary Curriculum (Majority of Classrooms)
High/Scope Curriculum
Creative Curriculum
Other/No Curriculum
PROGRAM LEVEL
CLASSROOM LEVEL
Teacher Backgrounds
Ethnicity (African-American, Hispanic)
Teacher Education and Experience
Years Teaching Overall (Head Start and other)
BA, AA only
Teacher salary (deviation
Tea
che
r
Atti
tud
Classroom Process Quality
ECERS-R Total Score
Quality Composite
Assessment Profile Individualizing
Assessment Profile Scheduling
Assessment Profile Individualizing
Quality of Language/Literacy Environment
ECERS-R Language Score
Teacher-Child Interactions
Arnett Caregiver Interaction Scale
Classroom Structural Quality
Child-Adult Ratio

The results of these analyses are summarized in Table 4.2.

**Predicting ECERS-R Total Score.** High quality classrooms are those where programs have a higher percentage of non-minority students, a higher percentage of language-minority students (e.g., Spanish speakers), and teachers who have greater knowledge about early childhood education practices.

**Predicting ECERS-R Language Scale.** Classrooms with higher quality of language activities and materials were those whose teachers had higher scores for attitudes and knowledge about early childhood education practice. Having a teacher with a BA or AA, or having a higher percentage of non-minority students or a higher percentage of language-minority students was associated with higher quality at the trend level, but did not reach statistical significance.

**Predicting Assessment Profile Individualizing score.** Family income was significant and teacher salary was a non-significant trend in explaining the variation in teacher individualizing practices. Classrooms with a focus on individualizing the classroom activities for individual students (an indicator of greater quality) are located in Head Start programs with greater numbers of families with relatively higher incomes.

**Predicting Teacher Arnett Caregiver Interaction Scale score.** Curriculum, teacher experience, and teacher attitudes and knowledge were significant predictors, with the strongest being the teachers' attitudes and knowledge of early childhood education practice, followed by programs that used the Creative Curriculum. Programs that used the High/Scope Curriculum had classrooms with somewhat higher scores for teacher sensitivity, although this did not reach statistically significant levels. Classrooms with sensitive and responsive teachers are those that use the Creative Curriculum, where the teachers have more years of teaching experience, and where the teachers hold more positive attitudes and knowledge about practices in early childhood education.

The teachers' education level, indicated by whether teachers had an associate's or bachelor's degree or not, was not significantly related to teacher sensitivity, even though it was related in earlier analyses without program level factors included. These results further suggest that the relationship between teacher education and classroom quality is not direct, but rather is mediated by their knowledge and attitudes towards early childhood education practice as well as by the type of curriculum used and the teacher's level of experience.

**Predicting Quality Composite Score.** Classrooms with higher scores for quality on this indicator (comprising the ECERS-R Language Scale, and the Assessment Profile Scheduling and Learning Environment Scales) were those from Head Start programs with a higher percentage of language-minority students, and whose teachers had more positive attitudes and knowledge of early childhood education practice.

Table 4.2 Summary of Program- and Classroom-Level Factors Predicting Classroom Quality, Fall 2000.

| | Predicting ECERS-R Total (Intercept=4.8) | | Predicting AP Individualizing (Intercept=3.5) | | Predicting Caregiver Interaction Scale (Intercept = 70.96 | | Predicting Child-Adult Ratio (Intercept= 5.4) | | Predicting ECERS R Language (Intercept=4.80) | | Predicting Quality Composite Score (Intercept=.003) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | Signif[1] | Estimate | Signif[1] | Estimate | Signif[1] | Estimate | Signif[1] | Estimate | Signif[1] | Estimate | Signif[1] |
| **Program-Level Factors** | | | | | | | | | | | | |
| Percent parents with some college or higher | 0.55 | | 0.52 | | 10.59 | | -2.57 | | 0.4645 | | 0.4629 | 0.06 |
| Percent parents earning $1500/month or more | 0.44 | | 3.82 | **0.01** | 2.83 | | 0.50 | | 0.3601 | | 0.1883 | |
| Percent non-minority students | 0.82 | **0.01** | 0.14 | | 3.24 | | -0.87 | | 0.6781 | 0.06 | 0.4254 | 0.06 |
| Percent language-minority students | 1.04 | **0.02** | 0.22 | | -1.98 | | -1.31 | | 0.8444 | 0.10 | 1.2704 | **0.01** |
| High Scope Curriculum | 0.30 | | -0.40 | | 5.33 | 0.10 | -0.88 | | 0.2408 | | 0.2171 | |
| Creative Curriculum | 0.33 | | -0.38 | | 6.41 | **0.02** | -0.56 | | 0.2676 | | 0.2639 | |
| Average Annual Teacher Salary | 0.10 | | -0.34 | 0.10 | 1.62 | | -0.81 | **0.01** | 0.08252 | | -0.1082 | |
| **Classroom-Level Factors** | | | | | | | | | | | | |
| Teacher having BA or AA | -0.09 | | 0.10 | | -1.09 | | -0.04 | | -0.2349 | 0.09 | -0.09462 | |
| Teacher Attitudes & Knowledge | 0.11 | **0.00** | 0.04 | | 1.55 | **0.00** | 0.04 | | 0.159 | **0.00** | 0.1363 | **0.00** |
| Years Teaching Experience | 0.01 | | 0.01 | | 0.17 | **0.05** | 0.00 | | 0.01154 | | 0.006333 | |
| African-American Teacher | -0.14 | | 0.06 | | -1.75 | | 0.14 | | -0.2208 | | -0.04489 | |
| Hispanic Teacher | -0.18 | | 0.04 | | -3.91 | | 0.21 | | -0.3437 | | -0.2133 | |
| Teacher Salary (deviation from program mean) | 0.03 | | 0.15 | | 1.09 | | 0.04 | | 0.1074 | | -0.03496 | |

*1. Significance levels in bold indicate statistically significant effects at $p < .05$.*

**Predicting Child-Adult Ratio.** Even after controlling for other factors within the classroom, high quality classrooms in terms of lower child-adult ratios are those where teachers are paid more.

**Summary of Multi-Level Model Results.** Overall the results support the model and enhance our understanding of factors that explain quality in Head Start. The results suggest that Head Start programs that provide for a common integrative curriculum across classrooms and that pay their teachers well have sufficient resources available to positively influence classroom quality, through the quality of teachers hired, and their experience and attitudes and knowledge. The factors included in the model do a good job of explaining the more "process-oriented" aspects of quality, such as learning materials, quality of language activities, and teacher-child interactions, but they do less well at predicting the more structural aspects of quality, such as child-adult ratios.

## G. Summary

Head Start classrooms continue to show good levels of quality, based on the indicators of quality measured in FACES 2000. These levels of quality are consistent from the first cohort, and the consistency is evident across a wide variety of the indicators. Head Start teachers are qualified and experienced (although as a group they do not have the same level of credentials as public school teachers), and there appear to be substantially more teachers with higher educational attainment in this cohort compared with the first FACES cohort, in 1997.

The role of teacher attitudes and knowledge, experience and education has been illuminated by these analyses. In general, when all three factors are included, the direct relationships occur most consistently and strongly for attitudes and knowledge about early childhood education practice, and to a lesser extent teacher experience, rather than with teacher education. When these other factors were not included we found a significant relationship between teacher education and classroom quality, suggesting that the role of teacher education in influencing classroom quality is an indirect one. Teachers with higher levels of education have more positive attitudes and knowledge about early childhood education practice, and they are more likely to be in classrooms rated higher in quality. Thus, teacher attitudes and knowledge mediates the role of teacher education in explaining classroom quality.

Interestingly, programs using an integrated curriculum (such as High/Scope and Creative Curriculum) also have teachers with positive attitudes and knowledge about early childhood education practice. Both of these factors appear to have the strongest effect on teacher sensitivity and responsiveness compared with other indicators of quality.

The multi-level results suggest that variations in the quality of Head Start classrooms may be explained by characteristics of the families and children they serve, by the curriculum used in the program, and by teacher attitudes and knowledge about early childhood education practice. The results suggest that Head Start classroom quality may be affected by factors beyond the classroom door, that are characteristics of the program and the families who participate.

## KEY FINDINGS

1. Head Start quality has been observed to be consistently good, over time, using a variety of indicators;
2. Head Start teachers have lower teaching qualifications compared with pre-k teachers in public elementary schools, but as a group they have become more qualified in 2000, with more of them having a graduate school degree;
3. Head Start teachers in 2000 are also younger, compared with those in 1997-1998, and more of them have been teaching in Head Start for two years or less. These newer teachers are also the ones most likely to have a graduate school degree;
4. teacher backgrounds, qualifications and experience are related to their attitudes and knowledge of early child development practices;
5. classrooms with higher levels of quality are those whose teachers have higher levels of education, experience, and knowledge and attitudes of early childhood education practices;
6. the relationship between teacher education and classroom quality is explained by teacher's attitudes and knowledge of early childhood education practices, so that teachers who are more educated have more positive attitudes and knowledge, which translates into higher levels of classroom quality; and
7. factors at the program level, including curriculum use, teacher salaries, and parent demographics, are also related to observed quality in Head Start classrooms and to the characteristics of teachers placed in these classrooms.

## REFERENCES

Abbott-Shim, M. & Sibley, A. (1998). Assessment Profile for Early Childhood Programs: Research version. Atlanta, GA: Quality Assist.

Abbott-Shim, M, Lambert, R, & McCarty, F. (2000). Structural model of Head Start classroom quality. *Early Childhood Research Quarterly*, 15, 1, 115-134.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? Journal of Applied Developmental Psychology, 10, 541-552.

Blau, D. (1997). The production of quality in child care centers. *Journal of Human Resources*, 32, 354-387.

Bredekamp, S. (1986). The reliability and validity of the early childhood classroom observation scale for accrediting early childhood programs. *Early Childhood Research Quarterly*, 1, 103-118.

Burts, D.C., C.H. Hart, R. Charlesworth, & L. Kirk. (1990). A comparison of frequencies of stress behaviors observed in kindergarten children in classrooms with developmentally appropriate versus developmentally inappropriate practices. *Early Childhood Research Quarterly*, 5, 407-423.

Cost, Quality & Child Outcomes Study Team. (1995). *Cost, quality and child outcomes in child care centers, public report, second edition.* Denver: Economics Department, University of Colorado at Denver.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale: Revised Edition.* New York: Teachers College Press.

Hayes, C., Palmer, J., & Zaslow, M. (1990*). Who cares for America's children? Child care policy for the 1990s.* Report of the Panel on Child Care Policy, Committee on Child Development Research and Public Policy, National Research Council. Washington, DC: National Academy Press.

Howes, C., Phillips, D., & Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based child care. *Child Development.* 63. 449-460.

Layzer, J., Goodson., B., & Moss, M. (1993). *Final report volume 1: Life in preschool.* Cambridge, MA: Abt Associates, Inc.

Phillips, D. & Howes, C. (1987). *Indicators of quality in child care: Review of the research.* In D. A. Phillips (Ed.), Quality in child care: What does research tell us? (pp. 1-19). Research Monograph of the National Association for the Education of Young Children, Vol. 1. Washington, DC: NAEYC.

Phillips, D., Mekos, D., Scarr, S., McCartney, K. & Abbott-Shim, M. (2000). Within and beyond the classroom door: Assessing quality in child care centers. *Early Childhood Research Quarterly,* 15, 4, 475-496.

Resnick, G., McKey, R.H., & Klayman, D. (2001). *The evaluation of early childhood education programming in the 30 Abbott School Districts: First-year report on program implementation and descriptions of children and families.* (Contract No. A43078). Trenton, NJ: New Jersey Department of Human Services and New Jersey Department of Education.

Ruopp, R., Travers, J., Glantz, F., & Coelen, C. (1979*). Children at the center: Final results of the National Day Care Study.* Boston: Abt Associates.

Scarr, S., & Eisenberg, M. (1993). Child care research: Issues, perspectives and results. Annual Review of Psychology, 44, 613-644.

Scarr, S., M. Eisenberg, & K. Deater-Deckard. (1992). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, 9, 131-151.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24(4), 323-355.

Smith, T., Kleiner, A., Parsad, B., & Farris, E. (2002). *Prekindergarten in U.S. public schools: 2000–2001* (NCES 2003-019). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Whitebook, M., Howes, C., & Phillips, D. (1989). *Who Cares? Child care teachers and the quality of care in America* (Final report of the National Child Care Staffing Study). Oakland, CA: Child Care Employee Project.

# Chapter V: Relationship of Program and Classroom Characteristics to Children's Cognitive Gains and Social Development in Head Start

The Family and Child Experiences Survey that began in fall 2000 (FACES 2000) found that the quality of Head Start centers and classes was generally good as judged by observational instruments that are widely employed to assay the quality of early childhood learning environments. An earlier round of the study that began in fall 1997 (FACES 1997) had similar findings. Both longitudinal studies also found that children in Head Start made significant progress toward national averages in some areas of early academic knowledge and skill, notably vocabulary knowledge and early writing skills. But in other areas, notably letter recognition and early math skills, children held their own but did not draw closer to national averages during the Head Start year. Even in those areas where they made significant progress, they still entered kindergarten considerably behind their more advantaged peers.

Why was this the case? Variations across local Head Start programs--in the achievement gains that children make in Head Start and in the levels of skill and knowledge with which they leave the program–offer potential explanations. Variations in cognitive and social-emotional development could be related to differences in classroom quality. Or they could be associated with differences in the type of curriculum or instructional approach that the programs or classroom teachers employ.

If this were indeed the case, then the performance of the national Head Start program might be improved by encouraging more local programs to improve their quality or make use of curricula or instructional approaches found to be associated with greater gains in children's knowledge and skills or greater improvements in their classroom conduct and social-emotional well-being.

This chapter explores variations in child achievement and behavior across local Head Start programs and classes. It uses multilevel modeling to test hypotheses about early education program and classroom characteristics that many child development scholars believe to be associated with enhanced cognitive growth or emotional maturation in preschool children (Phillips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000; Whitebook, Howes, & Phillips, 1989). The characteristics that are thought to make a difference for program effectiveness include the following:

- using an integrated and comprehensive preschool curriculum;
- having more ample program resources;
- providing classrooms that are of higher quality as early learning environments;
- employing a better prepared teaching staff;
- providing preschool services for a longer period each day;
- conducting educational activities in smaller groups with more personal adult attention to the needs and preferences of individual children; and
- encouraging parents to engage in more educational activities with their children at home.

This chapter examines relationships between these characteristics and several measures of children's cognitive development and classroom behavior, while controlling for the influence of other variables. Cognitive dependent variables consisted primarily of direct assessment measures of children's letter recognition and pre-reading skills, vocabulary knowledge, early writing skills, and early math skills. Dependent variables in the social-emotional realm consisted of teacher and parent ratings of children's approaches to learning tasks and cooperative and problem behavior. Control variables included measures of the socioeconomic and ethnic composition of the families and children participating in each program and classroom. Other control variables were characteristics of the child like age, sex, and disability status and measures of parents' literacy skills.

FACES 2000 included a wider range of program and classroom characteristics that could be related to differences in achievement than did FACES 1997. FACES 2000 added interview questions and observational procedures that looked more systematically at areas like the curriculum used in each center and classroom, training and support for that curriculum, teacher salary levels, teacher knowledge and beliefs, and the use of child portfolios and other procedures aimed at individualizing instruction. In addition, the sample design was modified to yield a larger number of sample children in each sample classroom. This produced more stable estimates of class means and more variation in child characteristics in each classroom subsample. The modified design made it possible to carry out multilevel regression analysis at three levels: the program, classroom, and child level. In FACES 1997, multilevel regression analyses could be carried out at only the center and child levels.

The multilevel analyses of FACES 2000 data did indeed show that some of the program and classroom level characteristics listed above were significantly related to variations in the size of the gains children made in Head Start. In presenting the analysis findings, we enumerate the program and classroom characteristics that seemed to make a difference. We describe the nature and size of the relationships involved. We also list factors that were not significantly related to gains in achievement or behavior and discuss possible reasons why hypothesized relationships failed to materialize.

## CONCEPTUAL FRAMEWORK

The conceptual framework that guided our analyses was a multi-level, multi-causal model of the influences that shape children's cognitive and social-emotional development and the factors that help determine the nature of the experience children have in Head Start. (See Figure 5.1.) This view posits that children's development in the early years is primarily a function of the experiences they have in their families. Children from low-income families, whose parents tend to have lower educational attainments than other parents, often do not experience the same extent or quality of intellectual stimulation at home as children from middle-class families (Phillips et al., 1998). Furthermore, their parents are less able to purchase high-quality supplementary or substitute care in the marketplace. In some cases, children from low-income families may also not receive as much emotional support from parents as they need for optimal development. A center-based early childhood learning environment such as Head Start may help provide experiences that would be beneficial for the development of all children, but especially for those from higher-risk family environments (NICHD Early Child Care Research Network, 2000). Furthermore, parents' involvement in their children's educational experiences may also be an important factor, and one that programs can foster.

The nature of the learning environment that a given child experiences in Head Start depends on the training and experience of teachers in the program, and the resources available to them in terms of facilities, materials and teaching assistants. Programs with more resources are likely to be better able to provide adequate facilities and materials and recruit and retain talented and well-prepared teachers (Whitebook, Howes, & Phillips, 1989).

But the character of the classroom environment that a given Head Start program is able to provide for a child is not just a question of program resources. It also depends on the educational philosophy to which the program adheres, and the kind of curriculum centers and teachers are encouraged to follow. Other things equal, children would be expected to do better in programs that employ well-thought-out curricula that are comprehensive and integrated in terms of educational activities and assessment methods. This is especially the case if the program is able to provide teachers with adequate training and support in the curriculum. At the same time, children's progress in a given cognitive or social-emotional area depends on whether the program's basic philosophy and curriculum of choice are supportive of efforts to bolster that area of child development.

## RESEARCH QUESTIONS

The analyses reported in this chapter used multilevel regression models to address the following research questions:

1. Do Head Start programs and classes differ in the average achievement levels that children have attained when they leave the program? Do they differ in the cognitive gains children make during the program year? Do they differ in the extent of changes children show in their cooperative social behavior or conduct problems?

2. Do children in programs that employ one of the two integrated curricula that are widely used in Head Start – Creative Curriculum or High/Scope – show larger cognitive gains or behavioral improvements than children in programs that employ other curricula?

3. Do children in programs that have more ample resources, as indicated by paying higher salaries to their lead teachers, show larger cognitive gains or behavioral improvements than children in programs that have less ample resources?

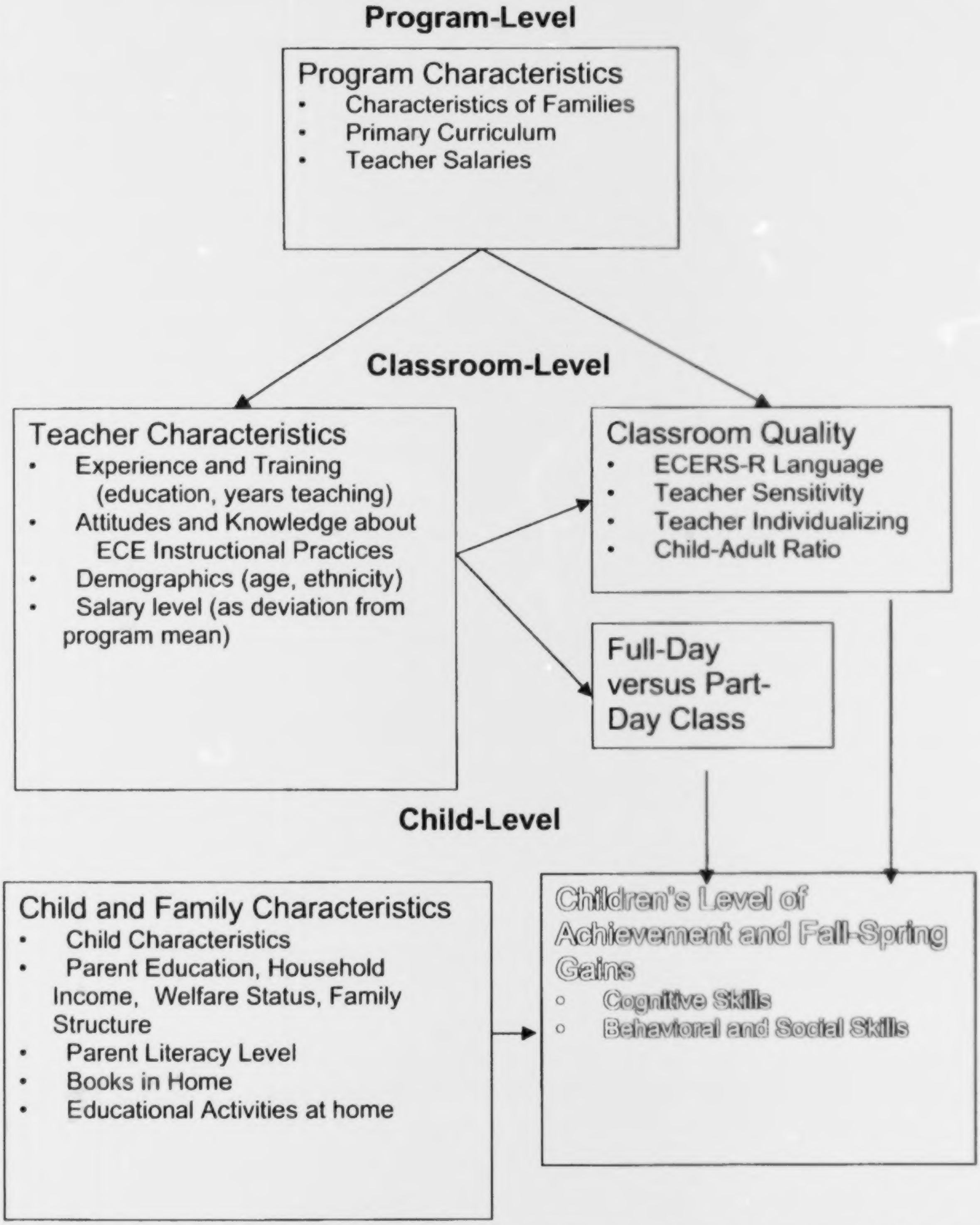

Program-Level
Program Characteristics
Characteristics of Families
Primary Curriculum
Teacher Salaries
Classroom-Level
Teacher Characteristics
Experience and Training (education, years teaching)
Attitudes and Knowledge about ECE Instructional Practices
Demographics (age, ethnicity)
Salary level (as deviation from program mean)
Classroom Quality
ECERS-R Language
Teacher Sensitivity
Teacher Individualizing
Child-Adult Ratio
Full-Day versus Part-Day Class
Child-Level
Child and Family Characteristics
Child Characteristics
Parent Education, Household Income, Welfare Status, Family Structure
Parent Literacy Level
Books in Home
Educational Activities at home
Children's Level of Achievement and Fall-Spring Gains
Cognitive Skills
Behavioral and Social Skills

Figure 5.1. Analytical Model of Multi-Level Factors Predicting to Classroom Quality and Children's Achievement and Gains in the Head Start Year

4. Do children in Head Start classes that are of higher quality, as indicated by their receiving higher scores on the Language scale of the Early Childhood Environment Ratings Scale – Revised (ECERS-R) or the Caregiver Interaction Scale (CIS), show larger cognitive gains or behavioral improvements than children in classes that are of lower quality on these measures?

5. Do children in Head Start classes led by better prepared teachers show larger cognitive gains or behavioral improvements than children in classes led by teachers who are less well prepared? Indicators of teacher preparation that were examined included whether the teacher had a Bachelors' Degree or Associates' Degree, her years of teaching experience, her annual salary as a deviation from the program mean salary, and her score on a scale that measured positive attitudes and knowledge about early childhood educational practices.

6. Do children who participate in Head Start classes for a longer period each day – who attend "full-day" classes – show larger cognitive gains or behavioral improvements than children who participate for a shorter period of time – who attend "part-day" classes?

7. Do children in Head Start classes with lower Child:Staff ratios, and that provide more attention to the needs and preferences of individual children, as indicated by a higher score on the Assessment Profile Individualizing scale, show larger cognitive gains or behavioral improvements than children who are in classes with higher Child:Staff ratios or lower Individualizing scores?

8. Do children whose parents do more educational activities at home with their children show larger cognitive gains or behavioral improvements than children whose parents do fewer educational activities? We examined whether children whose parents read to them on a daily basis at home showed greater gains in Head Start than children whose parents read to them less often. Frequency of reading was reported by the parents themselves in parent interviews in the fall of 2001.

## ANALYSIS METHOD

The analysis method used to examine associations between Head Start program and class characteristics and children's cognitive and social-emotional development was multilevel linear regression modeling, using the SAS PROC MIXED computer program (Singer, 1998; Bryk & Raudenbush, 1992). Multilevel modeling shows how the average achievement scores of a sample of classes, schools, or other educational units (in the present case, Head Start programs and classes) relate to a set of characteristics of those units, such as measures of program demographics and classroom quality. Simultaneously, this type of modelling can examine how the achievement scores of individual children in each program and class relate to a set of child-level characteristics, such as child demographics and home literacy activities. The method provides a numerical estimate of how sizable the program-to-program and class-to-class variation in average scores is, relative to the child-to-child variation in scores within classes.

The primary dependent variables were the gains each child made between the fall and the spring of the Head Start year in their cognitive assessment or behavior ratings scale scores. Models were also constructed of assessment and ratings scores attained by Head Start children in the FACES national sample in the fall of 2000 and the spring of 2001.[21] Each analytic model had three levels. The first level involved variation in average assessment scores or average gain scores across the 43 programs in the FACES national sample, expressed as deviations of the program means from the overall mean score for the entire sample. The second level involved variations of class means from the overall program means. And the third level involved variation in individual children's scores or gain scores around the class means.

> There were three levels of independent variables used to model or predict the assessment scores or gain scores. At the program level, the independent variables consisted of measures that represented the curriculum employed by the program, average teacher salary levels in the program, and average demographic and socioeconomic characteristics of the children who attended each program and their families. At the classroom level, independent variables consisted of measures of teacher preparation, teacher background characteristics, whether the class was of full-day or part-day duration, and indicators of classroom quality such as the ECERS-R Language scale and Caregiver Interaction Scale. Class-level variables also included measures of the demographic composition of the class, expressed as deviations from the average demographic characteristics of the program. At the child level, the independent variables were measures that represented demographic characteristics of the child; socioeconomic, cultural, and structural characteristics of the family; parent literacy levels; disability status of the child; and the frequency of parental reading to the child.

Statistical tests were made as to whether a given set of independent variables (program-level, class-level, and child-level) improved the model's fit to the data, over and above simpler models that did not include that set. Tests were also done as to whether the regression coefficient for a given independent variable was reliably greater than zero. Details about variable definitions, means and ranges, reliability of measures, and statistical tests used to ascertain the reliability of findings are described in the Appendix of the report.

## FINDINGS

Multilevel regression analyses of the assessment and ratings measures showed that there were significant relationships between some of the program and class characteristics identified above and variations in children's cognitive or behavioral gains. The following factors seemed to make a difference for children's progress:

---

[21] In the multilevel regression modeling, assessment scores were converted to "W-ability scores," based on IRT scaling of item difficulties carried out by the test developers. These scale scores are purported to have equal-interval properties that are desirable in regression modeling, particularly of gain scores. In other analyses, standard score versions of the assessment scores were used. These scores show how Head Start children performed compared to national norms. But they do not have as strong equal-interval properties as the W-ability scores.

1. programs using an integrated curriculum, particularly the High/Scope Curriculum;
2. programs having higher teacher salaries;
3. teachers having Bachelors' or Associates' Degrees;
4. children attending full-day rather than part-day classes; and,
5. parents reporting that they read to their children more frequently.

None of these factors was related to increased gains in all cognitive or behavioral indices examined. But most were related to gains on two or more outcome measures, and the relationships were in the hypothesized direction. The following sections describe the nature and size of these relationships.

### A. Use of an Integrated Curriculum

The High/Scope Curriculum is a comprehensive, integrated preschool curriculum that has a long history of research and development. It grew out of the Perry Preschool Project, an intensive, relatively small-scale intervention that was found to have long-term effects on children's achievement in a random-assignment evaluation study. It is the second most popular curriculum employed in Head Start programs, as described in Chapter III of this report. In the present analyses, children in programs that employed the High/Scope Curriculum were found to have made greater gains than children in programs that did not employ one of the two integrated curricula that are most widely used by local Head Start programs. (In the interests of simplicity, henceforth we shall refer to these curricula other than Creative Curriculum or High/Scope by the term "other curricula.") The greater gains were on measures in both the cognitive and social-emotional domains.

**Gains in pre-reading and oral communication skills.** Children in programs that employed the High/Scope Curriculum made small but significantly greater gains in letter recognition skills than children in programs that employed other curricula. In IRT scale-score terms, the average Head Start child made a gain on the Woodcock-Johnson-Revised Letter-Word Identification task of just under 10 scale points from fall to spring of the Head Start year. Children in programs employing the High/Scope Curriculum showed an average gain of 12.6 scale points on the WJ-R LWI from fall to spring ($p < .001$), whereas children in programs employing the Creative Curriculum or other curricula made gains of about 9 scale points ($p < .001$). The regression coefficient for the High/Scope Curriculum in the three-level regression analysis, which is an estimate of the difference in average gains between it and other curricula adjusted for the influence of related variables, was 3.66 ($p = .01$).

The other popular integrated curriculum, Creative Curriculum, had a positive coefficient in the regression analysis of Letter-Word Identification scores in the spring, but did not reach significance (2.66, $p < .10$). It also had a positive coefficient in the gain analysis, but one that was not significantly different from zero (1.92, n.s.).

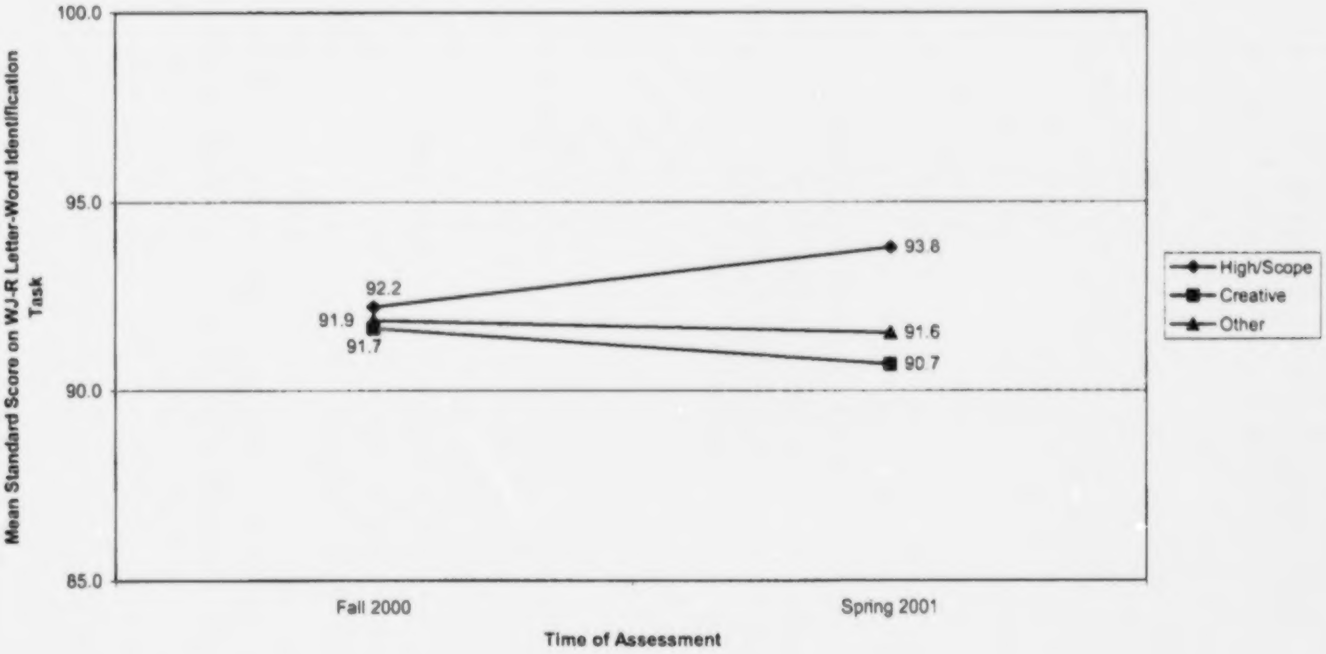
When differences were expressed in standard score terms, the High/Scope group showed a mean gain of 1.6 standard score points, going from a mean standard score of 92.2 in the fall to 93.8 in the spring. (Figure 5.2). By comparison, children in programs employing the other widely popular integrated curriculum, the Creative Curriculum or other curricula held their own against national norms.

Children in programs using the High/Scope Curriculum were found to make greater gains as well on a criterion-referenced measure of oral communication skills. This difference was only at the trend level, however. The "Social Awareness" measure assessed children's ability to tell an adult basic information about themselves such as their age, and month and year of birth. The regression coefficient for the High/Scope Curriculum showed a gain 0.21 points greater than that for other curricula ($p = .098$).

**Improvement in cooperative behavior.** The use of the an integrated curriculum was found to be associated with greater gains for children in the social-emotional realm. Children in programs employing the High/Scope Curriculum showed larger gains in cooperative classroom behavior from fall to spring of the program year than children in programs employing other curricula. They also showed more pronounced declines in hyperactive behavior from fall to spring.

Children in programs following the High/Scope Curriculum showed a mean increase of 2.3 points on the Cooperative Classroom Behavior rating scale completed by Head Start teachers (a change equivalent to .48 of a standard deviation). They went from a mean score of 14.5 in the fall to a mean of 16.8 in the spring ($p<.001$). By comparison, children in programs following the Creative Curriculum showed an increase of 1.8 points ($p<.001$) and those in programs following other curricula increased by 1.7 points ($p<.001$) (.37 and .36 of a standard deviation, respectively). The regression coefficient for the High/Scope Curriculum in the three-level regression analysis, which is an estimate of the difference in average gains between it and other curricula adjusted for the influence of related variables, was 1.26 ($p < .05$).

**Figure 5.2 Children in Head Start Programs Using High/Scope Curriculum Show Greater Gains in Letter Recognition Skills**



**Decline in hyperactive behavior.** Children in programs following the High/Scope Curriculum exhibited significant improvement in their scores on a Total Behavior Problems rating scale completed by Head Start teachers (p = .03). In particular, they showed greater improvement on the Hyperactive subscale of the Problem Behavior rating scale. They showed a mean decline of 0.19 points (p < .10) on this scale, going from a mean rating of 1.39 in the fall to a mean of 1.20 in the spring (a change of .13 of a standard deviation). By comparison, children in programs following the Creative Curriculum showed a non-significant mean decline of -.08 points, while children in programs following other curricula showed a non-significant mean decline of -.13 points. The regression coefficient for the High/Scope Curriculum in the three-level regression analysis, which is an estimate of the difference in average declines between it and other curricula, adjusted for the influence of related variables, was –0.32 (p < .05). In the regression analysis of Total Behavior Problems, the coefficient for the High/Scope Curriculum had a value of –1.19 (p = .01).

## B. Higher Teacher Salary Levels

We explored whether children attending Head Start programs with higher average teacher salary levels would make greater progress in their cognitive and social-emotional development. The multivariate analyses showed this to be the case with respect to children's pre-reading and oral communication skills, and their cooperative and problem behavior in the classroom.

**Gains in pre-reading and oral communication skills.** Average Annual Salary for Lead Teachers was associated with greater gains in letter recognition. The regression coefficient for Mean Teacher Salary Level in the three-level regression analysis of gains, which is an estimate of the difference in LWI scale scores associated with each $10,000 increment in average teacher salaries, adjusted for the influence of related variables, was 1.96 (p = .009). In standard score terms, the highest teacher salary group (top quartile) showed a gain of less than one standard score point. Children in programs with lower average teacher salary levels (bottom three quartiles) showed a slight and non-significant decline in their standard scores.

Children in programs with higher average teacher salaries made greater gains as well on the criterion-referenced "Social Awareness" measure. The regression coefficient for the Mean Teacher Salary Level showed an increased gain of 0.18 points for every $10,000 increment in mean salary (p = .008).

**Improvement in cooperative behavior.** Higher teacher salaries were found to be associated with greater gains for children in the social-emotional realm. Children in programs with higher average teacher salary levels showed larger gains in cooperative classroom behavior from fall to spring of the program year than children in programs with lower teacher salary levels. They also showed more pronounced declines in hyperactive behavior from fall to spring.

Children in programs in the highest quartile of teacher salaries showed a mean increase of 3.1 points on the Cooperative Classroom Behavior rating scale completed by Head Start teachers (a change equivalent of .65 of a standard deviation). They went from a mean score of 14.6 in the fall to a mean of 17.7 in the spring ($p<.001$). By comparison, children in programs in the middle two quartiles of teacher salary showed an increase of 1.7 points ($p<.001$), while those in programs in the lowest quartile increased by 1.5 points ($p<.001$) (equivalent to .34 and .30 of a standard deviation, respectively). (Figure 5.3.) The regression coefficient for Mean Teacher Salary Level in the three-level regression analysis, which is an estimate of the increase in average gains for every $10,000 increment in mean salary, adjusted for the influence of related variables, was 1.18 ($p < .001$).

**Decline in hyperactive behavior.** Children in programs with higher teacher salary levels exhibited significant improvement in their scores on the Hyperactive subscale of the Problem Behavior rating scale completed by Head Start teachers. Children in programs in the highest quartile on teacher salary levels showed a mean decline of 0.35 point ($p = .013$) on this scale, going from a mean rating of 1.39 in the fall to a mean of 1.04 in the spring (an effect size of .23 of a standard deviation). By comparison, children in programs in the middle two quartiles on teacher salary level showed no change (-0.12 point), while children in programs in the lowest quartile also showed no change (-.01 point). (Figure 5.4.) The regression coefficient for Mean Teacher Salary Level in the three-level regression analysis, which is an estimate of the difference in average declines for every $10,000 increment in mean teacher salary, adjusted for the influence of related variables, was –0.18 point ($p < .05$).

### C. Teachers With Bachelors' or Associates' Degrees

The possession of a four-year college degree or an Associates' Degree in education or a closely-related field is among the most widely accepted indicators of teacher preparation. One of the current performance goals of the national Head Start program is to have all local programs staffed by teachers of whom a majority have Bachelors' Degrees or Associates' Degrees. We explored whether the lead teacher having a BA or AA degree made a difference in children's progress on cognitive or social measures.

Figure 5.3 Children in Head Start Programs With Higher Teacher Salaries Show Larger Gains in Cooperative Classroom Behavior

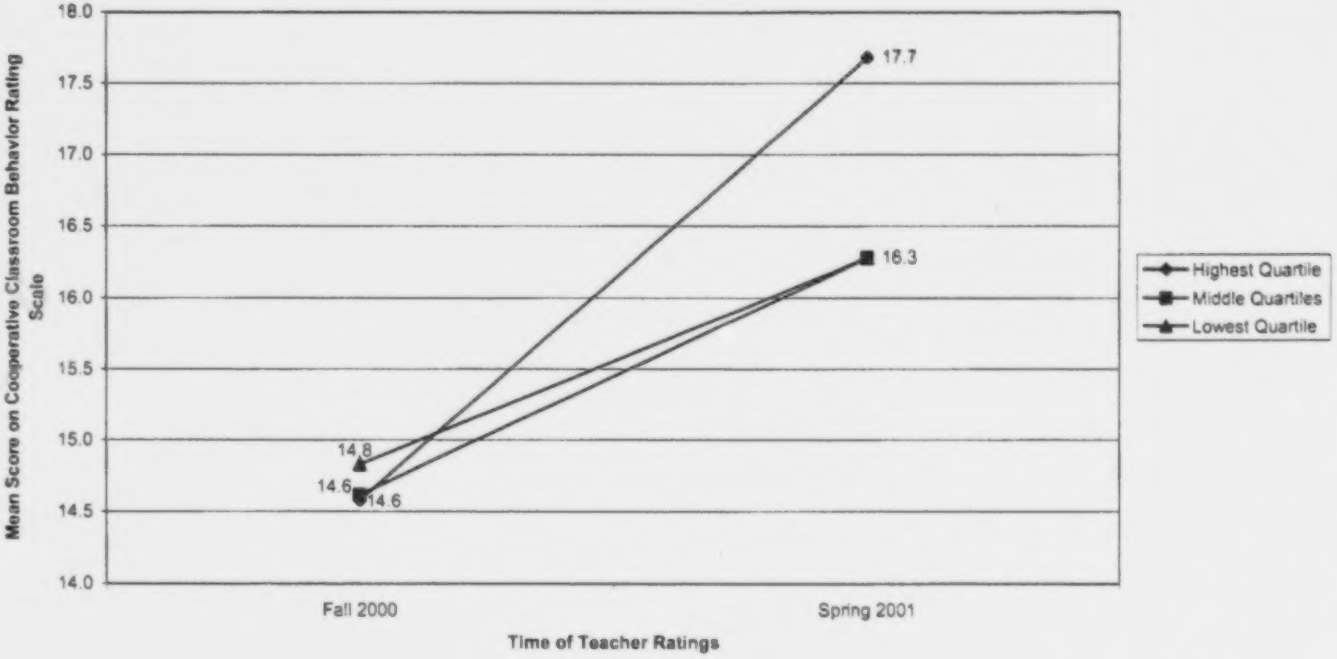
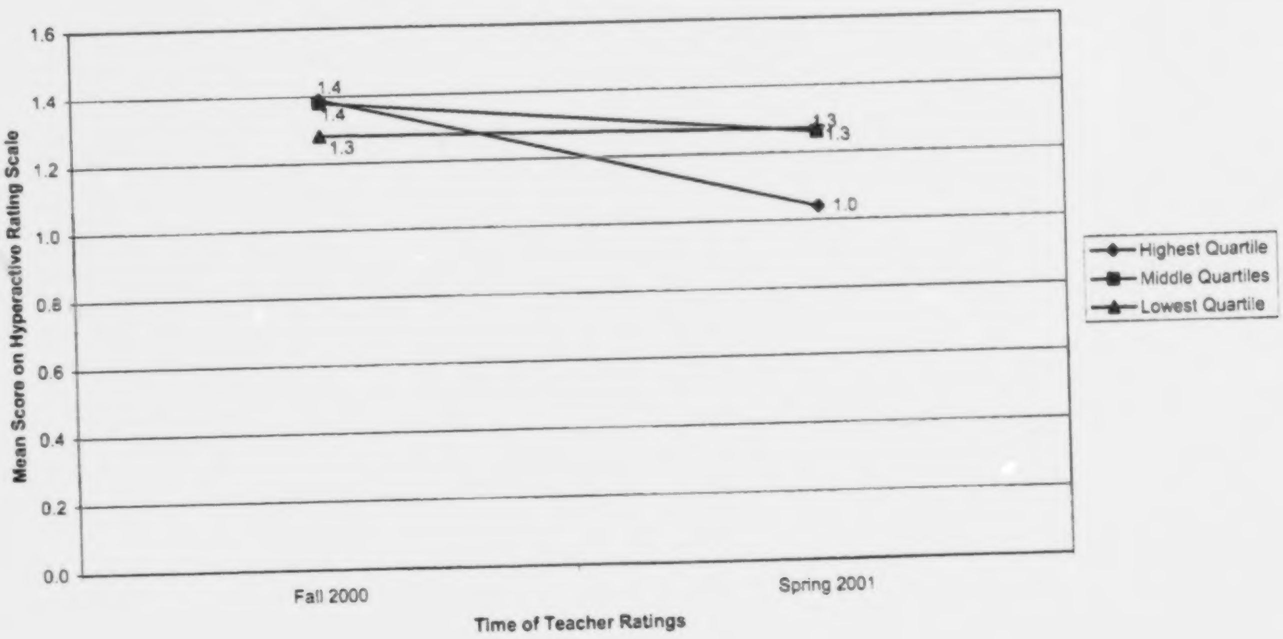


Figure 5.4 Children in Head Start Programs With Higher Teacher Salaries Show Larger Declines in Hyperactive Behavior

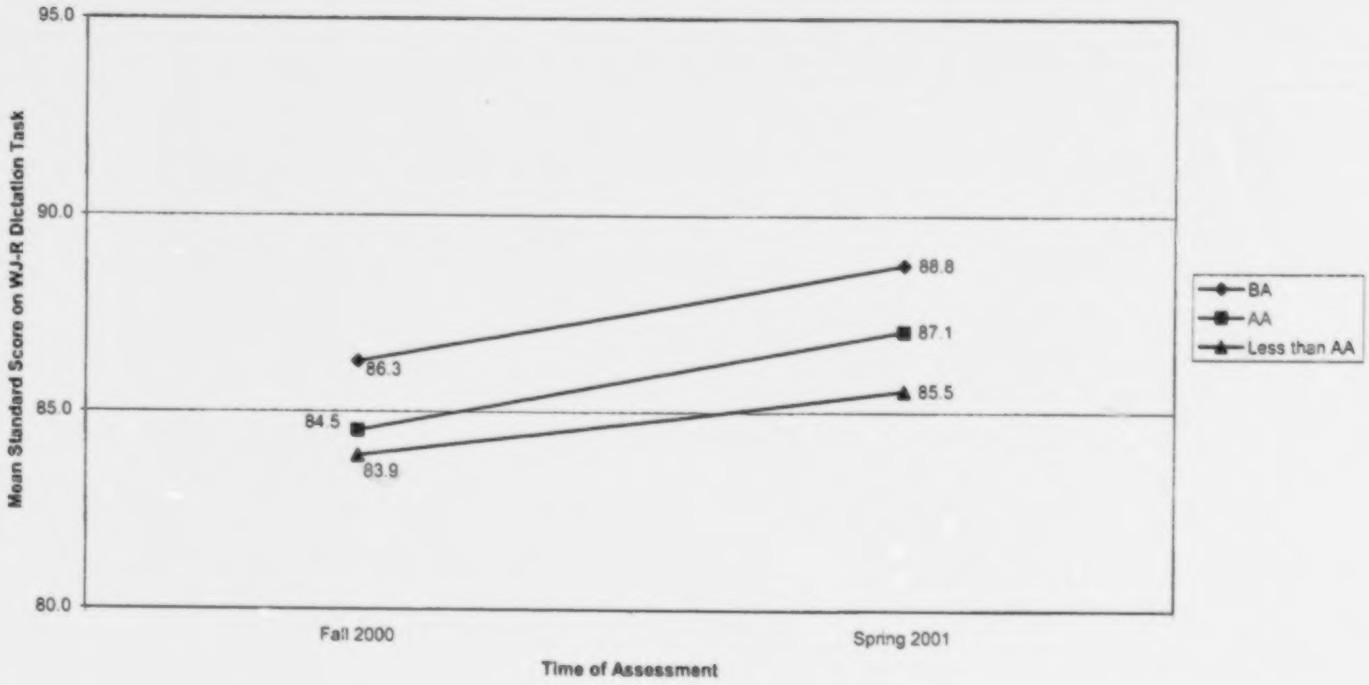


On several of the cognitive assessment measures, children in classes taught by teachers with BA or AA degrees ended the program year with mean scores that were higher than those of children in classes taught by teachers with less than an AA degree. However, these children had also had higher mean scores at the beginning of the year. This may reflect a situation in which Head Start programs that hire teachers with college credentials tend to serve families with higher parent education and income levels than are typical for Head Start nationwide. What was less certain was whether the children taught by teachers with higher educational credentials made greater gains from fall to spring than children taught by teachers with lesser credentials. Early writing skills was one cognitive area in which there was evidence of greater gains as well as higher achievement levels. However, the evidence was not unambiguous.

**Gains in early writing skills.** In standard score terms, and without adjustment for the effects of related variables, the picture was reasonably clear. Children in Head Start classes taught by lead teachers with Bachelors' Degrees or Associates' Degrees had higher mean scores on the Woodcock Johnson-Revised Dictation task in the fall – 86.3 and 84.5, respectively — than children in classes taught by teachers with less than an Associates' Degree, who had a mean score of 83.9. But the children in the former classes also made significant gains toward national averages, whereas children in the latter group did not. The gains were 2.48 standard score points ($p = .03$) for children whose teachers had Bachelors' Degrees; 2.55 standard score points ($p = .03$) for children whose teachers had Associates Degrees; and 1.67 standard score points ($p = .21$) for children whose teachers had less than an Associates' Degree. (Figure 5.5.) The respective gains represented effect sizes of .18, .19, and .12 of a standard deviation.

In the three-level analysis of spring Dictation scores, children in classes taught by teachers with BA's or AA's had a significant regression coefficient of 6.14 IRT scale points ($p = .01$). This meant that children in these classes had a mean score in the spring that much higher than the mean for children in classes taught by teachers without those credentials. In the fall analysis, children in classes taught by teachers with BA's or AA's had had a regression coefficient that was also significant, though apparently smaller (5.46 scale points, $p = .009$), which meant that they started with higher scores in the fall. The class-level variables as a set did not improve model fit in the fall, whereas they did in the spring. However, in the multilevel analysis of gains on the Dictation task, the regression coefficient for full-day classes (0.63) was not significant ($p = .77$).

Figure 5.5 Children in Head Start Classes Taught by Teachers With Bachelors' or Associates' Degrees Show Gains in Early Writing Skills



## D. Full-Day Versus Part-Day Classes

As of the 2000-2001 school year, the majority of children who attended Head Start participated in part-day classes that were conducted in morning or afternoon sessions only. We explored whether children benefited more from the program in terms of academic achievement if they attended full-day classes. Children in FACES 2000 who did attend full-day Head Start programs made greater gains in several areas than children who attended part-day.

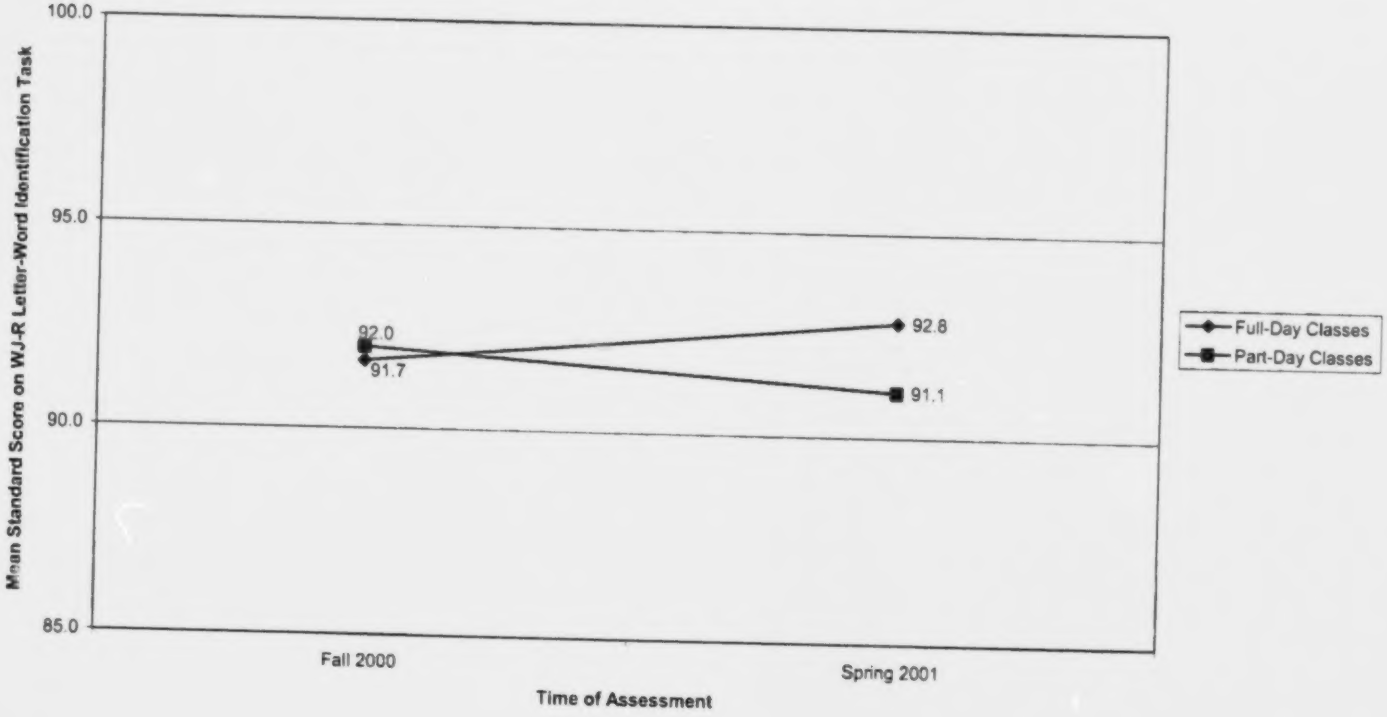
**Gains in pre-reading and early writing skills.** Children in full-day Head Start classes made larger gains in letter recognition skills than children in part-day classes. Children in full-day Head Start classes showed a mean gain on the Woodcock-Johnson Revised Letter-Word Identification task of 12 points in IRT scale-score terms ($p < .001$). Children in part-day classes showed a mean gain of 8.7 points ($p < .001$). The regression coefficient for full-day classes in the three-level regression analysis, which is an estimate of the difference between these classes and part-day classes adjusted for the influence of related variables, was only significant at the trend level (1.81, $p = .067$).

In standard score terms, the full-day group showed an average gain on the WJ-R LWI task of 1.2 standard score points ($p = .06$), whereas the part-day group merely held their own against national averages, showing a non-significant decline of 0.9 points ($p = .12$). (Figure 5.6.)

Children in full-day Head Start classes made greater gains as well in early writing skills, although the statistical evidence here was more ambiguous. In standard score terms, and without adjustments for the effects of related variables, children in full-day classes had a mean gain from fall to spring of 3.5 standard score points ($p = .004$) on the Woodcock-Johnson Revised Dictation task. They went from a mean of 84.8 in the fall to a mean of 88.3 in the spring. The gain was equivalent to an effect size of .25. By contrast, children in part-day classes went from a mean of 85.0 in the fall to a mean of 86.1 in the spring, a non-significant difference of 1.1 standard score points ($p = .162$). (Figure 5.7.)

In the three-level analysis of spring Dictation scores, children in full-day classes had a significant regression coefficient of 7.80 IRT scale points ($p = .005$). This meant that children in full-day classes had a mean score in the spring that much higher than the mean for children in part-day classes, with related factors controlled. In the fall analysis, children in full-day classes had had a regression coefficient that was only marginally higher (4.10 scale points, $p = .086$). Furthermore, the class-level variables as a whole did not improve model fit in the fall, whereas they did in the spring. However, in the multilevel analysis of gains on the Dictation task, the regression coefficient for full-day classes (3.85) was not significant ($p = .129$).

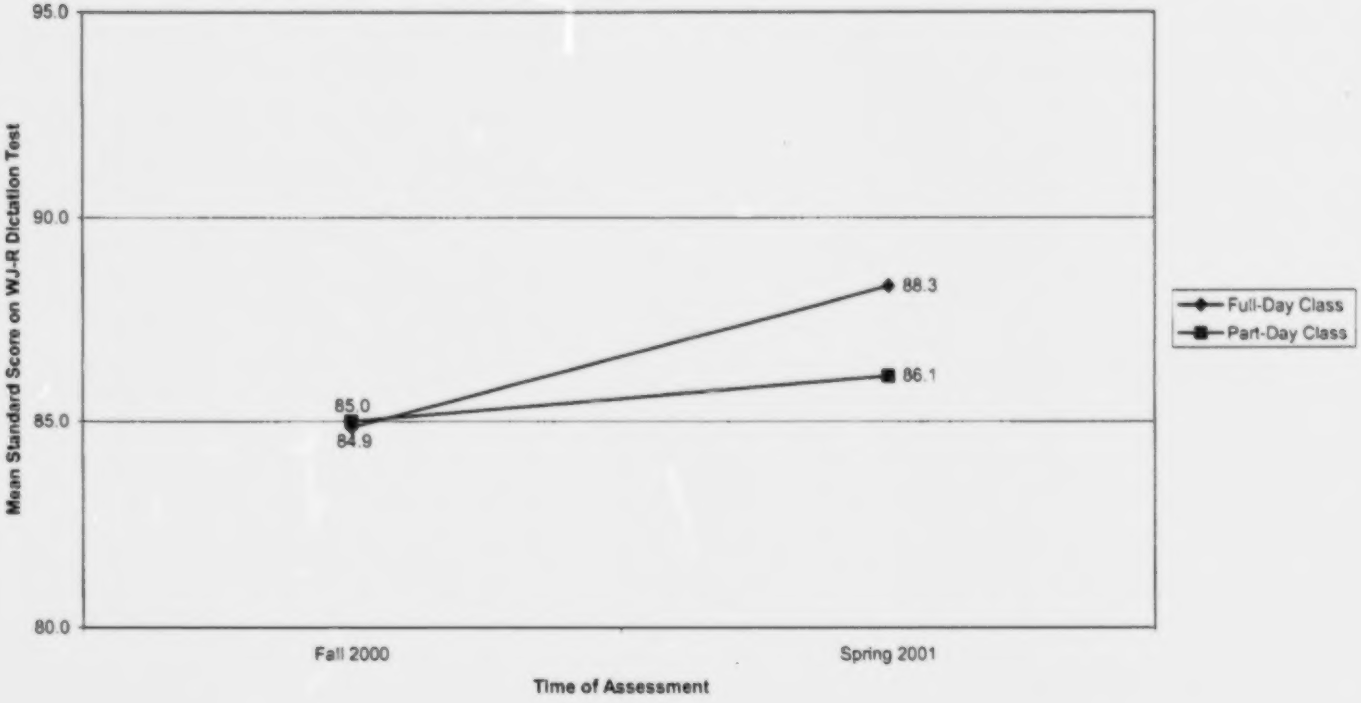
Figure 5.6 Head Start Children in Full-Day Classes Show Larger Gains in Letter Recognition Skills Than Those in Part-Day Classes



## E. More Frequent Parental Reading to Children

Children are in preschool programs for only a limited time, both in terms of hours of each day and months out of the child's life. However, preschool programs may extend their influence by encouraging parents to engage in more frequent and more effective educational activities at home with their children. The national Head Start program recognized the importance of this function by stating, in its performance measures framework, that one of the major objectives of the program is, to "strengthen parents as the primary nurturers of their children." Therefore, we decided to consider frequency of parental reading as an additional variable of interest in considering children's outcomes. Analyses showed that more frequent parental reading in the fall was associated not only with higher initial achievement for children as they entered the program, but also with larger gains during the program year. Larger gains were observed both in vocabulary knowledge and letter recognition skills.

**Figure 5.7 Children in Full-Day Head Start Classes Show Greater Gains in Early Writing Skills**



**Gains in vocabulary knowledge.** Parents were asked whether they read to their children, "not at all," "once or twice," "three to six times," or "every day" during the previous week. Parental responses to the question were entered into the three-level regression analysis as a set of dichotomous variables, with the most frequent response, "three to six times," as the omitted reference category. The reading responses were entered as child-level independent variables.

In the regression analysis of fall vocabulary test scores, children whose parents reported reading to their children "not at all" or "once or twice" had significantly lower mean scores than children whose parents reported reading "three to six" times. The mean score for children whose parents said they read "every day" was not significantly different from that of the "three to six times" group. In terms of IRT scale scores on the Peabody Picture Vocabulary Test, Third Edition, the mean for the "not at all" group was 1.93 points lower ($p < .05$), and the mean for the "once or twice" group was 1.83 points lower ($p < .001$), than the mean for the "three to six times" reference group.

In the regression analysis of the spring PPVT-III scores, the "not at all" and "once or twice" groups again had mean scores that were significantly lower (by 1.83 scale points, $p < .05$ and 1.35 points, $p < .01$, respectively) than that for the "three to six times" reference group. But now the "every day" group had a mean score that was significantly higher than the reference group mean (by 1.17 scale points, $p < .01$). Thus, there was a 3 scale-point difference between the vocabulary means of the highest and lowest reading groups. In the regression analysis of vocabulary gains, the "every day" group had a larger gain than the reference group (by 0.68 scale points), although the difference was only significant at the trend level ($p = .093$).

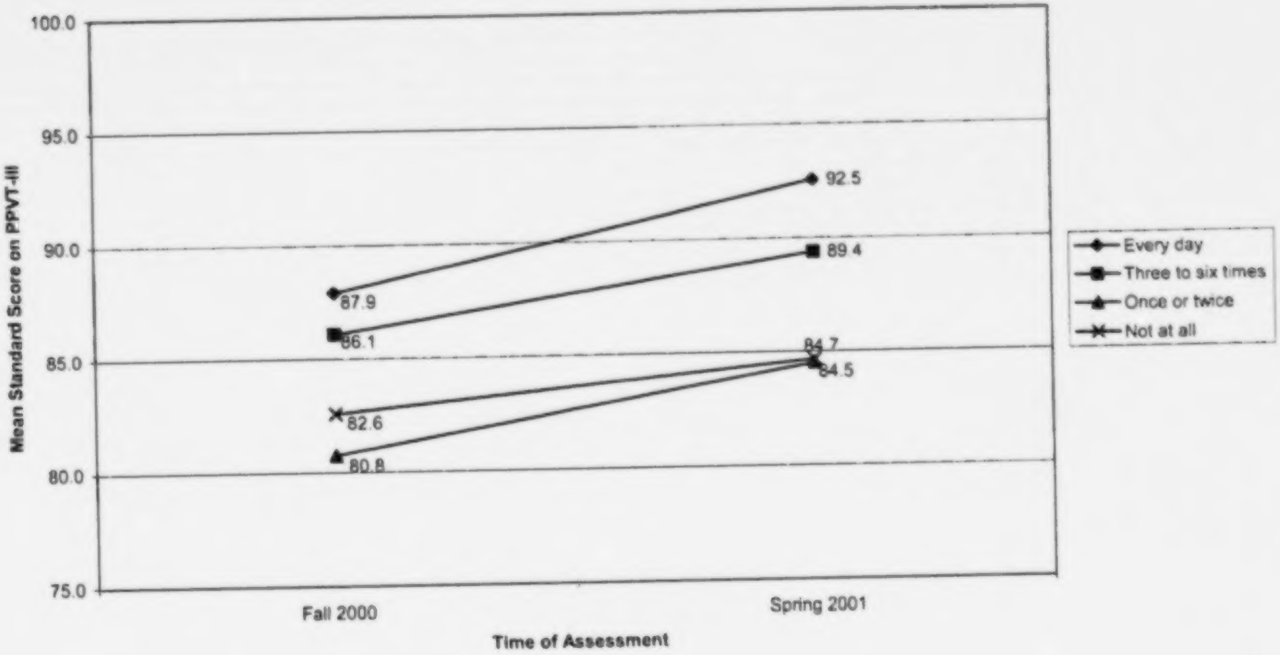
These differences related to frequency of parental reading were obtained even after controlling for parent education level, the mother's score on a measure of adult literacy (the K-FAST), and an indicator of the presence of books in the home. These measures were also entered into the regression analyses as child-level variables, and all were significantly related to children's vocabulary test scores in the fall and spring of the program year. None was related to the size of fall-spring gains in vocabulary scores, however.

The picture was similar, though not identical, when looked at in terms of mean standard scores for the parental reading groups without adjustments for the effects of related variables. (Figure 5.8.) All four groups showed significant gains in their vocabulary standard scores from fall to spring of the Head Start year. But the gain was smallest for the group whose parents said they read to the child "not at all" in the previous week (2.1 standard score points, $p < .05$). And the gain was largest for the group whose parents said they read to the child "every day" (4.6 standard score points, $p < .001$). However, the gain for the group whose parents read only "once or twice" was also sizable (3.8 standard score points, $p < .001$). When the gains were seen in terms of effect sizes, they ranged from .14 to .32 of a standard deviation, with the "every day" reading group having the largest effect size.

**Gains in pre-reading skills.** The parental reading groups showed differences in the gains children made on the Woodcock-Johnson Revised Letter-Word Identification task. Children whose parents reported reading to them only once or twice a week or less did not make as large gains in letter recognition skills as children whose parents reported reading to them three times a week or more. In the regression analysis of fall scores on the LWI test, the group means lined up in a fashion similar to that seen in the vocabulary analysis, but differences were not statistically significant. In the regression analysis of spring LWI scores, however, both the "not at all" reading group and the "once or twice" reading group had significantly lower means than the "three to six times" reference group. The respective differences, in terms of IRT scale scores, were –3.74 points ($p < .10$) and –2.95 points ($p < .05$). The "every day" reading group had a mean score that was not significantly different from that of the reference group.

Similar results were obtained in the regression analysis of fall-spring gains on the LWI task. Both the "not at all" reading group and the "once or twice" reading group had significantly smaller gains than the "three to six times' reference group. The respective differences in gains, in terms of IRT scale scores, were –3.23 points ($p < .10$) and –2.30 points ($p < .05$). The "every day" reading group had a mean gain that was not significantly different from that of the reference group. Again, these results controlled for the effects of parent education, parental literacy level, and the presence of books in the home.

Figure 5.8 Head Start Children Whose Parents Read To Them More Often Show Larger Vocabulary Gains



### F. Classroom Quality Indicators Not Found to Relate to Gains

The five program and class characteristics described above showed significant or marginally significant relationships with children's gains in Head Start. But there were two sets of characteristics that did not show the relationships with children's gains that were hypothesized. These were the indicators of classroom quality, and the indicators of child:staff ratio and individualized attention to the needs of each child.

As described in detail in Chapter IV and the Appendix, the FACES 2000 classroom observation battery contained a number of widely accepted indicators of the general quality of Head Start (or other preschool or childcare) environments. Two of these measures were chosen for inclusion in the three-level regression analyses of children's gains. One was the Language scale, a component of the Early Childhood Environment Rating Scale-Revised (ECERS-R). This component scale consists of a series of items observational items and ratings that deal with the frequency and quality of class activities related to oral language development, vocabulary building, and the nurturing of pre-reading and early writing skills. Thus, it seemed likely to relate to children's gains in these areas.

The second classroom quality indicator included in the three-level models was the Caregiver Interaction Scale, an observational rating measure of the emotional tone of teacher-child interaction and the lead teacher's sensitivity to children's needs and feelings. It seemed likely to relate to children's gains in the social-emotional domain. Neither of these expectations was supported by the FACES findings.

**ECERS-R Language scale.** Children in classrooms with higher ECERS-R Language scores had higher vocabulary test scores in the fall than children in classrooms with moderate or lower ECERS-R Language scores. In terms of standard scores, the mean score on the PPVT-III in the fall for children in Head Start classes in the highest quartile on the ECERS-R Language scale was 84.4. For children in classes in the middle two quartiles on the Language scale, the mean score was 80.5, while for children in classes in the lowest quartile, the mean score was 79.4. But children in higher, moderate, and lower quality classes all showed similar gains in vocabulary knowledge from fall to spring. Children in the highest quartile classes showed standard score gains of 4.26 points ($p < .001$, effect size of .25). Children in the middle two quartile classes showed standard score gains of 4.32 points, ($p < .001$, effect size of .25). And children in the lowest quartile classes showed standard score gains of 4.59 points ($p < .001$, effect size of .27). Thus, the same differences in vocabulary knowledge between higher and lower quality classes that were observed in the fall were still present in the spring. (Figure 5.9.) The higher quality language environments were not associated with larger gains in vocabulary knowledge.

Indeed, when relationships with other variables were controlled in the three-level regression analyses, even the differences in fall vocabulary scores associated with ECERS-R Language scores proved not statistically significant. This implies that the differences were better accounted for by the operation of other, related variables.

Higher ECERS-R Language scores were not associated with larger gains in letter recognition, early writing skills, or early math skills. Nor were they associated with improvements in children's cooperative or problem behavior in the classroom. Although the ECERS-R Language scale was the classroom quality indicator entered in the three-level regression models reported here, other correlation and regression analyses showed that the conclusions would not have been different had another quality indicator, such as the ECERS-R Total Score, been used instead. A Quality Factor weighted composite score was developed that incorporated the Assessment Profile Learning Environment and Scheduling scales as well as the ECERS-R Language scale. Analyses with that score produced non-confirmatory results as well.

**Caregiver Interaction Scale.** We hypothesized that the Caregiver Interaction Scale, an observation-based rating scale which reflects the sensitivity and emotional tone of teacher-child interaction, would be associated with improvements in children's cooperative and problem behavior in the classroom. This proved not to be the case. Higher CIS scores for the classroom teacher were associated with higher cooperative behavior ratings in the spring 2001 at only the trend level ($p < .10$). In the three-level analysis of gains in cooperative behavior, the coefficient for CIS scores had a value of zero. CIS scores were not significantly associated with reductions in Hyperactive Behavior or Total Behavior Problems either.

Higher CIS scores were not associated with greater gains in vocabulary, letter recognition, or early math skills. There was a significant positive association between higher CIS scores and higher mean scores on the WJ-R Dictation task in the spring of 2001. The regression coefficient signified a Dictation score that was .29 IRT scale points higher for each unit increase in the lead teacher's CIS score ($p = .03$). However, higher CIS scores were not significantly associated with greater fall-spring gains in this measure of early writing skills ($p = 20$).

**Child:Staff Ratio and More Individual Attention.** We hypothesized that lower child:adult ratios in Head Start classroom activities, and more attention to the needs and preferences of individual children would result in greater gains for children. These hypotheses were not confirmed by the data. Indeed, on some outcome measures, children actually showed greater gains in classrooms with more children per adult.

**Child:Staff Ratios.** The mean Child:Staff:Staff Ratio was a figure derived from counting the number of children in the Head Start classroom and dividing that number by the number of teachers or other adult staff members actively interacting with children. These counts were taken at two separate occasions on the day that the classroom was observed, and the two resulting ratios were averaged. When the mean Child:Staff Ratio was entered into the three-level regression analyses of children's cognitive gains as a class-level variable, it proved not to be significantly associated with gains in vocabulary knowledge, early writing, or early math skills. In the vocabulary analyses, the results were much like those for the ECERS-R Language score. That is, children in classes with lower, moderate, or higher Child:Staff Ratios all showed roughly equivalent gains in vocabulary knowledge from fall to spring of the Head Start year.

In the analyses of children's pre-reading skills, results were opposite to what was predicted. Children in classes with higher Child:Staff Ratios made significantly larger gains in letter recognition skills than those in classes with lower Ratios. In terms of standard scores, children in the highest quartile of classes on the Child:Staff Ratio showed an average gain of 1.2 standard score points on the WJ-R Letter-Word Identification task from fall to spring. They went from a mean standard score of 92.1 in the fall to a mean score of 93.3 in the spring. By comparison, children in classes in the middle two quartiles and lowest quartile merely held their own against national norms, showing non-significant declines of –0.6 standard score points (p = .28) and –0.9 standard score points (p = .55), respectively. (Figure 5.10.)

In the three-level regression analyses of LWI gains, the value of the regression coefficient for the mean Child:Staff Ratio was 0.57 (p < .05). This may be interpreted as the change in gain in LWI IRT scale scores that would be expected for every unit increase in the Child:Staff Ratio, net of the effects of other related variables. If the original hypothesis had been confirmed, this coefficient would be negative. Instead, it was reliably greater than zero in the positive direction.

Higher Child:Staff Ratios were also associated with behavioral gains. In the three-level regression analysis of gains in Cooperative Classroom Behavior, the coefficient for mean Child:Staff Ratio was 0.15 (p = .086). In the analysis of declines in the Total Behavior Problems, the coefficient for mean Child:Staff Ratio was –0.20 (p = .03). Both of these relationships were in the direction opposite to the expected one.

**Assessment Profile Individualizing Scale.** The Individualizing scale of the Assessment Profile instrument uses both observational and interview methods to assess the degree to which preschool teachers track the accomplishments of children in their classes and provide activities suited to the capabilities and interests of individual pupils. Class-level scores on this instrument did not relate to gains in any of the cognitive development areas. Nor did they show associations with improvements in the measures of social-emotional development.

**Figure 5.9 Children in Head Start Classes With Higher, Moderate, and Lower Quality Language Activities Show Parallel Gains in Vocabulary Knowledge**

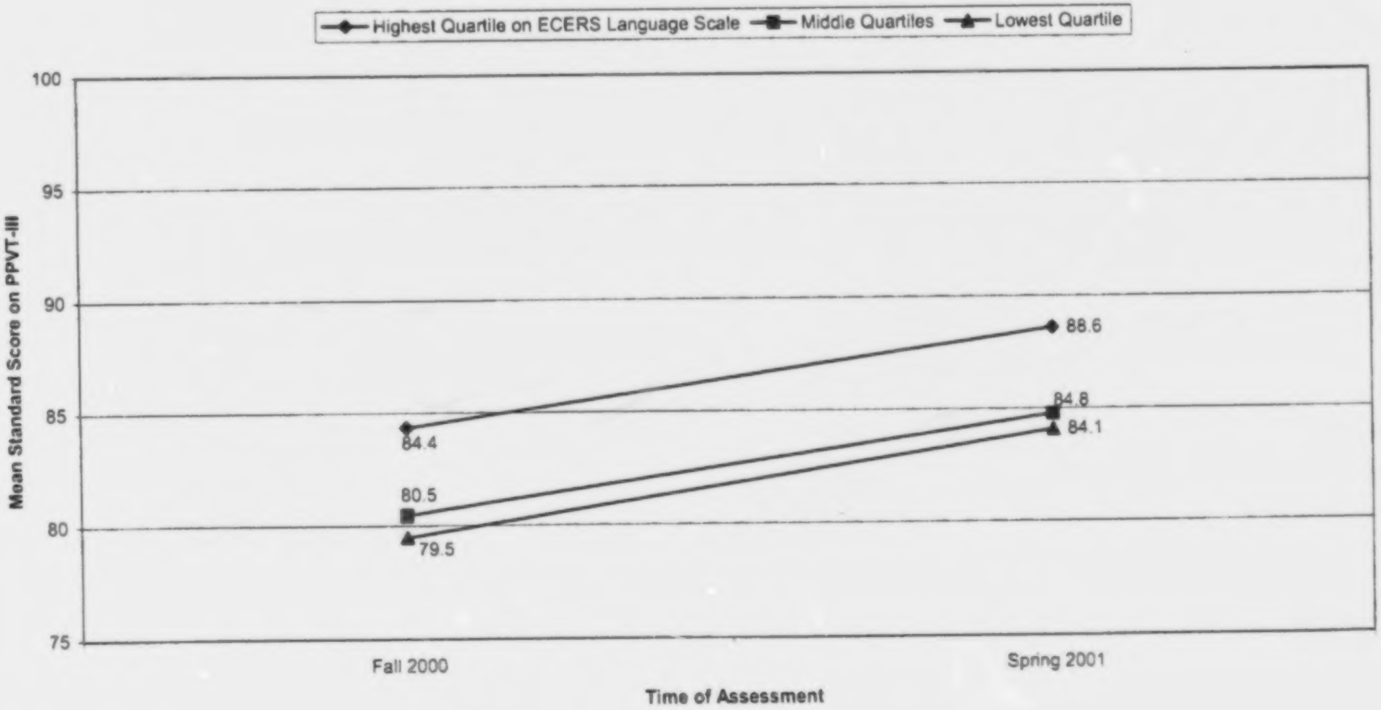
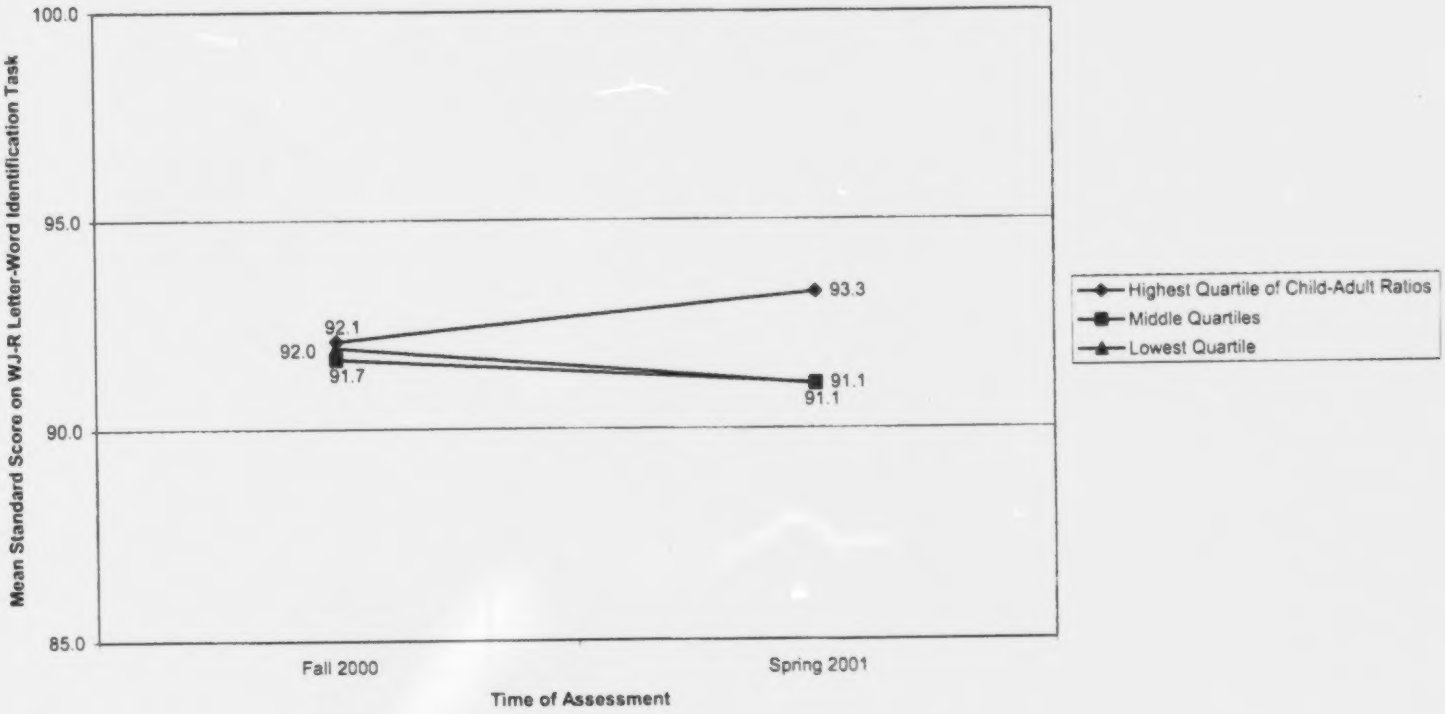


**Figure 5.10 Children in Head Start Classes With Higher Child-Adult Ratios Show Larger Gains in Letter Recognition**



## SUMMARY AND DISCUSSION

This chapter explored variations in child achievement and behavior across local Head Start programs and classes. It used multilevel modeling to test hypotheses about early education program and class characteristics that many child development scholars believe to be associated with enhanced cognitive growth or emotional maturation in preschool children. The conceptual framework posited that the gains a child makes in Head Start depends on the nature of the learning environment that he or she experiences in the local program. The nature of the learning environment depends in turn on the training and experience of teachers in the program and the resources available to them in terms of facilities, materials, and teaching assistants. Programs with more resources are likely to be better able to provide adequate facilities and materials and recruit and retain well-prepared teachers. Another hypothesis was that children would make larger gains in programs employing curricula that are comprehensive and integrated in terms of educational activities and assessment methods. Other expectations were that children would make more sizable gains in higher quality classrooms, in full-day as opposed to part-day classes, in classes with better child:staff ratios and more individualized attention to pupils, and in families where parents engaged in more educational activities with their children.

Analysis of longitudinal data from FACES 2000 showed that children's gains in Head Start were significantly related to several of the hypothesized characteristics of programs and classes. Specifically:

- Use of an integrated curriculum was linked to greater gains in several cognitive and social-emotional areas. Children in Head Start programs using High/Scope showed larger fall-spring gains in letter identification and cooperative classroom behaviors than children in programs using other curricula. Children in programs using High/Scope also showed greater improvement in total behavior problems and hyperactive problem behavior.

- Higher teacher salaries were linked to greater gains in several cognitive and social-emotional areas, including letter identification and cooperative classroom behavior. Children in programs with higher teacher salaries also showed greater improvement in hyperactive problem behavior during the Head Start year.

- Teachers' educational credentials were linked to greater gains in early writing skills. Children taught by Head Start teachers with Bachelors' Degrees or Associates' Degrees showed gains toward national averages in an assessment of early writing skills, whereas children taught by teachers with lesser credentials merely held their own against national norms.

- Provision of preschool services for a longer period each day was linked to greater cognitive gains. Children in full-day classes in Head Start showed larger fall-spring gains in letter recognition and early writing skills than did children in part-day classes.

- There was indirect evidence that encouraging parents to engage in more educational activities with their children at home could serve as a pathway to greater cognitive gains. Children whose parents reported reading to them every day showed larger fall-spring gains in vocabulary knowledge and letter recognition skills than children whose parents reported reading once or twice or less frequently per week.

Other analytic results were not in line with expectations. In particular:

- Within the generally good quality range of Head Start classrooms, variation in quality as measured by the ECERS-R Language scale or the Caregiver Interaction Scale was not associated with differences in fall-spring achievement gains across classes.

- Within the narrow range of child:staff ratios in Head Start, variation in child:staff ratios was not associated with or was negatively associated with differences in fall-spring achievement gains across classes.

The analysis results were generally supportive of the conceptual framework that the amount of resources available to a Head Start program and the curricular approach it uses can make a difference for children's progress in the program. The results also supported the notions that children could make greater gains if they had more exposure to comprehensive, integrated preschool activities. While we were not able to consider the program's effect on parental reading, it appears that benefits accrue from increased frequency of educational activities at home. At the same time, some provisos about the results should be noted. These include the following points:

1. Differences in cognitive gains, while statistically significant, were relatively modest in magnitude. By itself, each of the differences was not large enough to close the gap between where Head Start children typically end up at the end of the program year and the average achievement levels of American children at the start of elementary school. If several of the positive characteristics could be implemented simultaneously in a program, they might jointly make a more sizable difference, however.

2. Program- and classroom-related gains varied across cognitive areas. Significant gains were seen primarily with respect to letter recognition and early writing skills. The important areas of vocabulary and early math skills showed little variation in gains that was linked to specific program or class characteristics. Rather, children in programs and classes with different characteristics tended to show parallel gains (in vocabulary) or similar lack of gains (in early math skills).

3. Differences in achievement levels at the end of the Head Start year between children in Head Start programs with differing socioeconomic and ethnic composition were substantial. This was particularly the case with respect to vocabulary knowledge. The program and class characteristics studied here did little to narrow these gaps. This was partly because, as just noted, the studied characteristics were not linked to differential gains in vocabulary and math skills. But it was also because local

programs with higher average parent education and income levels tended to have more of the desirable program and class characteristics.

4. Findings of significant links between program and class characteristics and improvements in children's behavior have to be tempered by the realization that the measures of children's behavior made use of ratings by teachers and parents. Thus, it is possible that the observed relationships were partly due to differences in the rating patterns of different groups of teachers rather than (or as well as) to actual behavioral differences between groups of children.

5. The failure to find significant links between children's cognitive gains in Head Start and class-level scores on the ECERS-R Language scale may have to do with the generally good quality of Head Start classrooms and the limited range of variation in classroom quality that FACES found in its national samples of programs and classes (Peisner-Feinberg, Burchinal, Clifford, Culkin, Howes, Kagan & Yazejian, 2001). The same may be said of the failure to find significant links between children's improvements in the social-emotional domain and class-level scores on the Caregiver Interaction Scale. Studies encompassing broader ranges of quality of childcare and early education facilities have shown greater variations in classroom quality measures and significant relationships between quality measures and children's gains (Peisner-Feinberg & Burchinal, 1997; Bryant, Burchinal, Lau, & Sparling, 1994; NICHD Early Child Care Research Network, 2000; Phillips, McCartney, & Scarr, 1987; Whitebook, Howes & Phillips, 1989; but for another failure to find a relationship, see Kontos & Fiene, 1987).

At the same time, the FACES results should make us wary of claims that Head Start could produce dramatically larger achievement gains in children from low-income families simply by raising ECERS scores or other indicators of classroom quality. It may be that good classroom quality is a necessary but not sufficient condition for practically significant gains in specific cognitive or behavioral areas. It may be that further progress depends on discovering and applying instructional approaches that can bolster gains in specific areas. Preliminary findings from randomized intervention studies conducted in Head Start programs in New York state as part of the Head Start Quality Research Consortium studies suggest that children in Head Start can make strikingly larger gains in letter recognition and related skills with appropriate, research-based supplementary curricula (Fischel, Storch, Spira & Stoltz, 2003).

## REFERENCES

Bryant, D. M., Burchinal, M., Lau, L. B., and Sparling, J.J. (1994). Family and classroom correlates of Head Start children's developmental outcomes. *Early Childhood Research Quarterly, 9,* 289-309.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Cost, Quality & Child Outcomes Study Team. (1995). *Cost, quality and child outcomes in child care centers, public report, second edition.* Denver: Economics Department, University of Colorado at Denver.

Fischel, J., Storch, S.A., Spira, E.G., and Stoltz, B.M. (2003). Enhancing emergent literacy skills in Head Start: First year curriculum evaluation results. Poster presented at Society for Research in Child Development Biennial Meeting, Tampa Bay, Florida, April 26.

Howes, C., Phillips, D., and Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based child care. *Child Development.* 63. 449-460.

Kontos, S., and Fiene, R. (1987). Child care quality, compliance with regulations, and children's development: The Pennyslvania Study. In D. Phillips (ed.), *Quality in child care: What does the research tell us?* (pp. 57-79). Washington, DC: National Association for the Education of Young Children.

NICHD Early Child Care Research Network (ECCRN). (2000). The relation of child care to cognitive and language development. *Child Development, 71,* 823-839.

Peisner-Feinberg, E. S. and Burchinal, M. R. (1997). Relations between preschool children's child-care experiences and concurrent development: The Cost, Quality and Outcomes Study. *Merrill-Palmer Quarterly.* 43, 3. 451-477.

Peisner-Feinberg, E.S., Burchinal, M.R., Clifford, R.M., Culkin, M.L, Howes, C., Kagan, S.L., and Yazejian, N. (2001). The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Developmental Trajectories through Second Grade. *Child Development, 72(5),* 1534-1553.

Phillips, D., McCartney, K., and Scarr, S. (1987). Child care quality and children's social development. *Developmental Psychology, 23,* 537-543.

Phillips, D., Mekos, D., Scarr, S., McCartney, K. and Abbott-Shim, M. (2000). Within and beyond the classroom door: Assessing quality in child care centers. *Early Childhood Research Quarterly.* 15, 4. 475-496.

Phillips, M., Brooks-Gunn, J., Duncan, G.J., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the black-white test score gap. Pp. 103-145 in *The Black-White Test Score Gap*, edited by C. Jencks & M. Phillips, Washington DC: Brookings Institution Press.

Scarr, S., and Eisenberg, M. (1993). Child care research: Issues, perspectives and results. Annual Review of Psychology. 44. 613-644.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics,* 24(4), 323-355.

Whitebook, M., Howes, C., and Phillips, D. (1989). *Who Cares? Child care teachers and the quality of care in America* (Final report of the National Child Care Staffing Study). Oakland, CA: Child Care Employee Project.

**Table 5.1. Three-Level Regression Models of Assessment Scale Scores of Head Start Children in Fall and Spring of Program Year, and Fall-Spring Gains, 2000-2001**

| | Fall | | Spring | | Letter ID | |
|---|---|---|---|---|---|---|
| | Letter ID | | Letter ID | | Gains | |
| **Program-Level Predictor Variables** | *Unstandardized Regression Coefficients* | | | | | |
| High Scope Curriculum | 1.01 | | 4.48 | * | 3.66 | * |
| Creative Curriculum | 0.88 | | 2.66 | † | 1.92 | |
| Mean Teacher Salary Level | 1.21 | † | 3.10 | ** | 1.96 | ** |
| Proportion Non-Minority Children | -0.07 | | -0.38 | | -0.22 | |
| Program Mean Parent Education Level | 3.55 | ** | 4.76 | * | 1.20 | |
| Program Mean Family Income Level | -0.67 | | 0.02 | | 0.30 | |
| Proportion Language-Minority Children | 5.20 | | 3.07 | | -2.06 | |
| **Class-Level Predictor Variables** | | | | | | |
| Full-Day Class | 0.25 | | 1.99 | | 1.81 | † |
| Average ECERS Language Score | -0.49 | | -0.92 | | -0.41 | |
| Average Child-Adult Ratio | 0.34 | | 0.91 | ** | 0.57 | * |
| AP Individualizing Score | 0.68 | † | 0.37 | | -0.27 | |
| Average Lead Teacher Arnett Score | 0.03 | | 0.02 | | -0.01 | |
| Teacher BA or AA | 1.93 | * | 0.19 | | -1.71 | † |
| Years Teaching Experience | -0.02 | | 0.00 | | 0.02 | |
| Teacher DAP Beliefs Score | -0.36 | | -0.76 | * | -0.40 | |
| Black Teacher | -2.50 | * | -0.16 | | 2.33 | † |
| Hispanic Teacher | -0.13 | | -0.41 | | -0.36 | |
| Teacher Salary Deviation Score | -0.15 | | -1.55 | | -1.39 | † |
| Class Parent Education Level (deviation) | 1.62 | * | 2.72 | * | 1.09 | |
| Class Family Income Level (deviation) | -0.73 | | -0.42 | | 0.32 | |
| Class Proportion Non-minority (deviation) | 1.86 | | 5.23 | | 3.54 | |
| Proportion Language-Minority (deviation) | -1.59 | | 2.34 | | 3.84 | |
| **Child-Level Predictor Variables** | | | | | | |
| Parent Literacy Standard Score | 0.10 | ** | 0.06 | | -0.04 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| (KFAST) | | | | | | |
| Parent Education Level (deviation) | 0.98 | ** | 1.39 | ** | 0.39 | |
| Family Income Level (deviation) | 0.39 | | 0.66 | | 0.26 | |
| Welfare Status | -1.40 | | -1.51 | | -0.08 | |
| Books In Home | 1.50 | | 0.79 | | -0.72 | |
| Frequency of Reading to Child | | | | | | |
| Not At All | -0.55 | | -3.74 | † | -3.23 | † |
| One or twice | -0.68 | | -2.95 | * | -2.30 | * |
| Every day | 0.47 | | 0.65 | | 0.22 | |
| Age of Child in Months | 0.51 | *** | 0.68 | *** | 0.17 | † |
| Sex of Child | 0.46 | | 1.24 | | 0.79 | |

### Table 5.1. Three-Level Regression Models of Assessment Scale Scores of Head Start Children in Fall and Spring of Program Year, and Fall-Spring Gains, 2000-2001, continued

| | | | | | | |
|---|---|---|---|---|---|---|
| **Child-Level Predictor Variables** | | | | | | |
| Black Child | 1.77 | | 1.97 | | 0.29 | |
| Hispanic Child | -0.55 | | -1.92 | | -1.51 | |
| Language Minority Family | 1.07 | | -1.49 | | -2.45 | |
| Disability Status | -2.12 | * | -4.07 | ** | -2.00 | † |
| Mother-Father Family | -1.07 | | -0.49 | | 0.64 | |
| Neither Birth Parent In Home | 2.13 | | -1.55 | | -3.69 | † |
| *Continued* | | | | | | |
| **Intercept** | 301.22 | *** | 297.66 | *** | -2.90 | |
| Proportion of Variance Accounted For: | | | | | | |
| Between-Programs Variance | 100% | *** | 95% | *** | 100% | ** |
| Between-Classes Variance | 100% | *** | 92% | *** | 99% | ** |
| Within-Classes Variance | 7% | *** | 6% | *** | 1% | ** |
| | | | | | | |
| Total Variance | 15% | ** | 20% | *** | 12% | ** |
| | | | | | | |
| *** $p < .001$ | | | | | | |
| ** $p < .01$ | | | | | | |
| * $p < .05$ | | | | | | |
| † $p < .10$ | | | | | | |
| N = 957 | | | | | | |
| | | | | | | |

**Table 5.2. Three-Level Regression Models of Assessment Scale Scores of Head Start Children in Fall and Spring of Program year, and Fall-Spring Gains, 2000-2001**

| | Fall | | Spring | | Vocabulary | |
|---|---|---|---|---|---|---|
| | Vocabulary | | Vocabulary | | Gains | |
| **Program-Level Predictor Variables** | *Unstandardized Regression Coefficients* | | | | | |
| High Scope Curriculum | 0.38 | | 1.26 | | 0.91 | |
| Creative Curriculum | 0.88 | | 0.71 | | -0.15 | |
| Mean Teacher Salary Level | 0.57 | | 0.34 | | -0.17 | |
| Proportion Non-Minority Children | 4.65 | *** | 5.77 | *** | 1.25 | |
| Program Mean Parent Education Level | 2.22 | * | 2.59 | ** | 0.31 | |
| Program Mean Family Income Level | 0.39 | | -1.14 | | -1.61 | |
| Proportion Language-Minority Children | -1.11 | | 2.16 | | 3.01 | † |
| **Class-Level Predictor Variables** | | | | | | |
| Full-Day Class | -0.41 | | -0.36 | | 0.10 | |
| Average ECERS Language Score | 0.06 | | -0.13 | | -0.24 | |
| Average Child-Adult Ratio | 0.20 | | 0.20 | | 0.00 | |
| AP Individualizing Score | 0.17 | | 0.06 | | -0.04 | |
| Average Lead Teacher Arnett Score | -0.09 | | 0.03 | | 0.01 | |
| Teacher BA or AA | 0.07 | | -0.01 | | 0.00 | |
| Years Teaching Experience | -0.04 | | -0.05 | † | -0.02 | |
| Teacher DAP Beliefs Score | -0.05 | | -0.02 | | 0.02 | |
| Black Teacher | -0.28 | | -0.07 | | 0.04 | |
| Hispanic Teacher | 0.55 | | 0.73 | | 0.36 | |
| Teacher Salary Deviation Score | -0.25 | | -0.24 | | 0.04 | |
| Class Parent Education Level (deviation) | 1.64 | *** | 1.61 | *** | 0.02 | |
| Class Family Income Level (deviation) | -1.08 | | -0.51 | | 0.44 | |
| Class Proportion Non-minority (deviation) | 3.25 | * | 3.87 | ** | 0.69 | |
| Proportion Language-Minority (deviation) | -0.59 | | -2.46 | | -2.38 | |
| **Child-Level Predictor Variables** | | | | | | |
| Parent Literacy Standard Score (KFAST) | 0.13 | *** | 0.12 | *** | -0.02 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Parent Education Level (deviation) | 0.64 | *** | 0.65 | *** | 0.01 | |
| Family Income Level (deviation) | 0.20 | | 0.05 | | -0.13 | |
| Welfare Status | -0.66 | | -0.92 | † | -0.21 | |
| Books In Home | 1.63 | ** | 1.26 | ** | -0.31 | |
| Frequency of Reading to Child | | | | | | |
| Not At All | -1.93 | * | -1.83 | * | 0.12 | |
| One or twice | -1.83 | *** | -1.35 | ** | 0.44 | |
| Every day | 0.57 | | 1.17 | ** | 0.68 | † |
| Age of Child in Months | 0.81 | *** | 0.73 | *** | -0.08 | ** |
| Sex of Child | -0.23 | | 0.06 | | 0.32 | |

Table 5.2. Three-Level Regression Models of Assessment Scale Scores of Head Start Children in Fall and Spring of Program year, and Fall-Spring Gains, 2000-2001, continued

| | | | | | | |
|---|---|---|---|---|---|---|
| **Child-Level Predictor Variables** | | | | | | |
| Black Child | -3.31 | *** | -3.73 | *** | -0.33 | |
| Hispanic Child | -1.97 | *** | -2.62 | *** | -0.64 | |
| Language Minority Family | -8.51 | *** | -7.19 | *** | 1.32 | † |
| Disability Status | -2.47 | *** | -2.59 | *** | -0.13 | |
| Mother-Father Family | -0.65 | | -0.83 | * | -0.17 | |
| Neither Birth Parent In Home | 2.05 | * | 0.46 | | -1.52 | † |
| | | | | | | |
| **Intercept** | 0.55 | | 13.65 | ** | 13.23 | ** |
| Proportion of Variance Accounted For: | | | | | | |
| Between-Programs Variance | 99% | *** | 97% | *** | 56% | * |
| Between-Classes Variance | 72% | *** | 80% | *** | 21% | |
| Within-Classes Variance | 34% | *** | 33% | *** | 1% | * |
| | | | | | | |
| Total Variance | 53% | *** | 53% | *** | 3% | † |
| | | | | | | |
| ***$p < .001$ | | | | | | |
| **$p < .01$ | | | | | | |
| *$p < .05$ | | | | | | |
| †$p < .10$ | | | | | | |
| N = 1,984 | | | | | | |

**Table 5.3. Three-Level Regression Models of Behavior Rating Scores of Head Start Children in Fall and Spring of Program year, and Fall-Spring Gains, 2000-2001**

| | Fall Cooperative Behavior | | Spring Cooperative Behavior | | Cooperative Behavior Gains | |
|---|---|---|---|---|---|---|
| **Program-Level Predictor Variables** | *Unstandardized Regression Coefficients* | | | | | |
| High Scope Curriculum | -0.01 | | 1.30 | * | 1.26 | * |
| Creative Curriculum | -0.53 | | 0.06 | | 0.57 | |
| Mean Teacher Salary Level | -0.58 | † | 0.59 | * | 1.18 | *** |
| Proportion Non-Minority Children | -1.00 | | -0.47 | | 0.45 | |
| Program Mean Parent Education Level | 0.04 | | -0.20 | | -0.26 | |
| Program Mean Family Income Level | -0.80 | | -1.79 | * | -0.92 | |
| Proportion Language-Minority Children | -0.08 | | 0.37 | | 0.31 | |
| **Class-Level Predictor Variables** | | | | | | |
| Full-Day Class | -0.29 | | 0.13 | | 0.45 | |
| Average ECERS Language Score | -0.06 | | 0.04 | | 0.06 | |
| Average Child-Adult Ratio | 0.03 | | 0.17 | * | 0.15 | † |
| AP Individualizing Score | -0.12 | | -0.35 | * | -0.22 | |
| Average Lead Teacher Arnett Score | 0.03 | | 0.03 | † | 0.00 | |
| Teacher BA or AA | -0.10 | | -0.53 | | -0.43 | |
| Years Teaching Experience | -0.03 | | -0.02 | | 0.00 | |
| Teacher DAP Beliefs Score | -0.06 | | 0.01 | | 0.07 | |
| Black Teacher | -0.73 | | -0.96 | † | -0.33 | |
| Hispanic Teacher | -0.06 | | -0.39 | | -0.28 | |
| Teacher Salary Deviation Score | 0.23 | | -0.02 | | -0.24 | |
| Class Parent Education Level (deviation) | -0.44 | | -0.40 | | 0.07 | |
| Class Family Income Level (deviation) | 0.62 | | 0.10 | | -0.48 | |
| Class Proportion Non-minority (deviation) | 1.03 | | 1.16 | | -0.02 | |
| Proportion Language-Minority (deviation) | 0.34 | | -0.47 | | -0.79 | |
| **Child-Level Predictor** | | | | | | |

| **Variables** | | | | | | |
|---|---|---|---|---|---|---|
| Parent Literacy Standard Score (KFAST) | 0.00 | | 0.01 | * | 0.01 | |
| Parent Education Level (deviation) | 0.14 | * | 0.13 | † | -0.01 | |
| Family Income Level (deviation) | 0.22 | * | -0.01 | | -0.23 | * |
| Welfare Status | -0.27 | | -0.77 | *** | -0.50 | * |
| Books In Home | 0.40 | † | 0.07 | | -0.34 | |
| Frequency of Reading to Child | | | | | | |
| Not At All | -0.08 | | 0.22 | | 0.32 | |
| One or twice | -0.13 | | -0.14 | | -0.01 | |

**Table 5.3. Three-Level Regression Models of Behavior Rating Scores of Head Start Children in Fall and Spring of Program year, and Fall-Spring Gains, 2000-2001, continued**

| **Child-Level Predictor Variables** | | | | | | |
|---|---|---|---|---|---|---|
| Every day | 0.44 | * | 0.26 | | -0.17 | |
| Age of Child in Months | 0.22 | *** | 0.17 | *** | -0.05 | *** |
| Sex of Child | 1.31 | *** | 1.48 | *** | 0.18 | |
| Black Child | -0.02 | | -0.26 | | -0.24 | |
| Hispanic Child | 0.23 | | -0.18 | | -0.42 | |
| Language Minority Family | 0.08 | | 0.50 | | 0.40 | |
| Disability Status | -1.46 | *** | -0.99 | *** | 0.47 | † |
| Mother-Father Family | -0.11 | | 0.38 | * | 0.49 | * |
| Neither Birth Parent In Home | -0.54 | | 0.07 | | 0.60 | |
| | | | | | | |
| *Continued* | | | | | | |
| **Intercept** | 5.00 | | 6.79 | * | 1.66 | |
| Proportion of Variance Accounted For: | | | | | | |
| Between-Programs Variance | 77% | | 100% | | 65% | * |
| Between-Classes Variance | 11% | | 22% | * | 8% | |
| Within-Classes Variance | 14% | *** | 12% | *** | 2% | ** |
| | | | | | | |
| Total Variance | 16% | *** | 17% | *** | 7% | *** |
| | | | | | | |
| ***$p < .001$ | | | | | | |
| **$p < .01$ | | | | | | |
| *$p < .05$ | | | | | | |
| †$p < .10$ | | | | | | |
| N = 2138 | | | | | | |

**Table 5.4. Three-Level Models of Behavior Rating Scores of Head Start Children in Fall and Spring of Program Year, and Fall-Spring Gains, 2000-2001**

| | Fall Hyperactive Behavior | Spring Hyperactive Behavior | Hyperactive Behavior Declines |
|---|---|---|---|
| **Program-Level Predictor Variables** | *Unstandardized Regression Coefficients* | | |
| High Scope Curriculum | 0.08 | -0.23 | -0.32* |
| Creative Curriculum | 0.08 | -0.07 | -0.15 |
| Mean Teacher Salary Level | 0.08 | -0.11 | -0.18* |
| Proportion Non-Minority Children | -0.31 | -0.09 | 0.24 |
| Program Mean Parent Education Level | 0.12 | -0.33* | -0.45** |
| Program Mean Family Income Level | 0.00 | 0.47† | 0.47* |
| Proportion Language-Minority Children | 0.25 | 0.29 | 0.06 |
| **Class-Level Predictor Variables** | | | |
| Full-Day Class | 0.20† | 0.14 | -0.04 |
| Average ECERS Language Score | 0.00 | 0.06 | 0.06 |
| Average Child-Adult Ratio | -0.02 | -0.06* | -0.04 |
| AP Individualizing Score | 0.09† | 0.06 | -0.03 |
| Average Lead Teacher Arnett Score | -0.01† | -0.01 | 0.00 |
| Teacher BA or AA | -0.12 | 0.04 | 0.17† |
| Years Teaching Experience | 0.01† | 0.01 | 0.00 |
| Teacher DAP Beliefs Score | -0.02 | 0.01 | 0.03 |
| Black Teacher | -0.24 | -0.10 | 0.14 |
| Hispanic Teacher | -0.21 | 0.00 | 0.21 |
| Teacher Salary Deviation Score | 0.04 | 0.12 | 0.09 |
| Class Parent Education Level (deviation) | 0.14 | 0.17† | 0.03 |
| Class Family Income Level (deviation) | -0.03 | -0.03 | 0.00 |
| Class Proportion Non-minority (deviation) | -0.28 | -0.16 | 0.13 |
| Proportion Language-Minority (deviation) | 0.39 | 0.18 | -0.21 |
| **Child-Level Predictor Variables** | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Parent Literacy Standard Score (KFAST) | 0.00 | | 0.00 | | 0.00 | |
| Parent Education Level (deviation) | -0.06 | * | -0.06 | * | 0.00 | |
| Family Income Level (deviation) | -0.06 | † | -0.01 | | 0.05 | |
| Welfare Status | 0.01 | | 0.04 | | 0.03 | |
| Books In Home | -0.17 | * | -0.04 | | 0.14 | † |
| Frequency of Reading to Child | | | | | | |
| Not At All | -0.18 | | -0.01 | | 0.17 | |
| One or twice | -0.11 | | 0.11 | | 0.22 | ** |
| Every day | -0.19 | * | -0.10 | | 0.10 | |

**Table 5.4. Three-Level Models of Behavior Rating Scores of Head Start Children in Fall and Spring of Program Year, and Fall-Spring Gains, 2000-2001, continued**

| | | | | | | |
|---|---|---|---|---|---|---|
| **Child-Level Predictor Variables** | | | | | | |
| Age of Child in Months | -0.05 | *** | -0.03 | *** | 0.02 | *** |
| Sex of Child | -0.41 | *** | -0.53 | *** | -0.13 | * |
| Black Child | -0.21 | † | -0.05 | | 0.17 | |
| Hispanic Child | -0.12 | | -0.07 | | 0.07 | |
| Language Minority Family | -0.09 | | -0.20 | | -0.11 | |
| Disability Status | 0.62 | *** | 0.35 | *** | -0.26 | ** |
| Mother-Father Family | -0.07 | | -0.27 | *** | -0.19 | ** |
| Neither Birth Parent In Home | 0.02 | | 0.09 | | 0.07 | |
| | | | | | | |
| *Continued* | | | | | | |
| **Intercept** | 4.38 | *** | 3.84 | *** | -0.61 | |
| Proportion of Variance Accounted For: | | | | | | |
| Between-Programs Variance | 0% | | 100% | | 100% | * |
| Between-Classes Variance | 22% | | 16% | | 11% | |
| Within-Classes Variance | 10% | *** | 9% | *** | 3% | *** |
| | | | | | | |
| Total Variance | 11% | *** | 12% | *** | 6% | *** |
| | | | | | | |
| *** $p < .001$ | | | | | | |
| ** $p < .01$ | | | | | | |
| * $p < .05$ | | | | | | |
| † $p < .10$ | | | | | | |
| N = 2045 | | | | | | |

# Chapter VI: Relationship of Family and Parental Characteristics to Children's Cognitive and Social Development in Head Start

Head Start maintains a strong interest in supporting parents as the primary educators of their children. Considering this parental role, the FACES Technical Report I (ACYF, 2002)[22] highlighted the diversity of Head Start families as well as noting many of the challenges that they encountered, such as balancing work and child care responsibilities, and coping with maternal depression, exposure to violence, involvement with the criminal justice system, and substance use. It was also reported that many of these families possessed strengths and resilience in the face of such challenges, and that when it came to their children, they held many of the same fundamental hopes, goals, values, and beliefs that are commonly shared among families, particularly with regard to education.

This chapter will explore how parental and family characteristics are related to aspects of children's school readiness. FACES Technical Report I (ACYF, 2002) noted that parents held optimistic expectations for their children's early schooling and valued the long-term benefits of education. While the majority of parents' short-term goals for their children were focused on general academic skills, such as completing age-appropriate tasks and success in school, most also held specific long-term educational attainment goals for their children, such as graduating from high school and attending college. One essential theme that emerged was that parents felt a strong desire to have the best for their children and to instill values focused on education.

Despite facing various barriers to participation, Head Start families had a strong desire to be involved in their children's Head Start experience and valued their own involvement in the program. Although parents identified barriers to their participation in Head Start (primarily time constraints related to work or school), many still participated in Head Start activities. Parents felt it was important for them to participate because they felt their involvement helped their children, because their children seemed to enjoy it, or because it was meaningful to their children to have their parents participate in activities at their schools.

With this background information in mind, this chapter explores specific characteristics of families and parents that may be related to the development of school readiness in the current sample of Head Start children.

## RESEARCH QUESTIONS

In this chapter, the following research questions will be addressed:

1. How are family and parental characteristics related to children's cognitive and social development?
2. How are family interactions with children related to children's cognitive and social development?
3. How do parents interact with Head Start and how is their interaction related to children's cognitive and social development?

---

[22] Findings cited from the FACES Technical Report I include quantitative data from the main study, as well as qualitative data from the case study.

Using fall 2000 and spring 2001 FACES parent interview data, this chapter explores relationships between family and parental characteristics and child outcomes (including teacher ratings of social skills and behavior problems, and direct cognitive assessments). These data provide further insights into the development of cognitive and social skills, particularly with regard to family challenges and strengths. The family risk factors investigated in this chapter include maternal depression, exposure to violence, domestic violence, involvement in the criminal justice system, substance use, and cumulative social and economic risks. The chapter concludes with a presentation of protective family factors and how the Head Start experience may help families cope with multiple life challenges.

## FINDINGS

The findings presented in this chapter are primarily based on partial correlations, tests of group differences (t-test, ANOVA), and multivariate logistic regression models used to estimate levels of risk while controlling for demographic variables. Unless noted in the text, all findings tested significant ($p \leq .05$). Additional information on specific parent interview scales and child assessment measures is found in the Appendix. Means and standard deviations for scales from the parent interview and from the teachers' reports of children's behavior are also included in the Appendix.

### A. Prevalence of Family and Parental Risk Factors and Their Relationship to Children's Outcomes

**Maternal Depression.** Because depression is a frequent challenge facing low-income families with young children (Hall, Williams & Greenberg, 1985; Liaw & Brooks-Gunn, 1994), depression among the Head Start parents was measured using the CES-D Depression Scale (Radloff, 1975). Overall, parents had a mean score of 6.8 in the fall of 2000, which was in the mildly depressed range. While most parents were classified as not depressed (47.7 percent) or only mildly depressed (27.0 percent), one fourth of the parents (25.3 percent) were classified as moderately depressed (13.6 percent) or severely depressed (11.7 percent). From fall to spring, there was a small decline in the overall mean depression scores (spring 2001 mean score of 6.6), but the difference was not statistically significant.

The level of depression did not vary significantly by ethnicity. Similar proportions of African American parents (27.7 percent), Hispanic parents (27.4 percent), and White parents (25.0 percent) were classified as moderately or severely depressed.

Group differences, evaluated using t-tests, were found between parents who were moderately or severely depressed and those who were not at all or only mildly depressed. Parents who were moderately or severely depressed reported a lower household income, had a more external locus of control, had a more authoritarian parenting style (i.e., more directive and harsh), and spanked their children more frequently. When asked about activities with the children, parents who were moderately or severely depressed were more likely to report participating in fewer activities with their children and were also less likely to be involved with their children's Head Start program. A higher proportion of mothers living without a father in the home were classified as moderately or severely depressed (30.4 percent) than those who had a father present in the home (20.3 percent).

Results of partial correlations controlling for parental income, education, employment, and child gender, age, and ethnicity also revealed that the overall parental depression scale score was significantly correlated with parent and teacher reports of the children's behavior. Parents who were more depressed reported that children had more problem behavior, including being more aggressive, hyperactive, and withdrawn. Similarly, teachers also reported more aggressive, hyperactive, and overall problem behavior for children of parents who were more depressed. Negative correlations were also found between parental depression and children's positive social behavior ratings and emergent literacy, indicating that less depressed parents reported having children with better social and academic skills.

Significant correlations were also found between parental depression and selected child cognitive outcomes. Children of parents who were depressed did worse on one-to-one counting and early math tasks, as well as on the teacher reports of creativity (descriptions of the child outcome measures are found in the Appendix).

Additional risks associated with maternal depression were explored more fully through multivariate logistic regression models, controlling for mother's education and employment, household income, and child's race, age, and gender. Odds ratios and 95% confidence intervals for each model are presented in Table 6.2 (located at the end of the chapter). These risk estimates indicate that parents who were moderately or severely depressed, compared to parents who were not at all or only mildly depressed, were 1.5 times more likely to be single parents, and almost twice as likely to report that they, another household member, or a non-household biological parent had been arrested or charged with a crime since the birth of their Head Start children. In addition, depressed parents were 1.3 times more likely than non-depressed parents to drink or live with someone who drank alcohol, and 1.6 times more likely to have been exposed to violence in their neighborhoods or homes. The risk for screening positive for domestic violence greatly increased for those parents who were moderately or severely depressed. These parents were almost 3 times more likely to have been victims of domestic violence compared to parents who were only mildly depressed or not depressed.

**Exposure to Violence.** Neighborhoods have long been recognized in theory and research as important contexts for child development (Leventhal & Brooks-Gunn, 2000), but they can provide unique challenges when families are exposed to violence. Parents were asked about the violence they knew to occur in their neighborhoods, and were asked additional questions about their own personal exposure to violence and domestic violence, as well as their Head Start children's exposure to violence.

More than one fifth of all parents (22.5 percent) reported seeing nonviolent crime such as selling drugs or stealing in their neighborhoods in fall 2000 (15.3 percent more than once), as well as having been a witness to violent crime (22.4 percent; 17.2 percent more than once). Approximately 17.0 percent of the parents knew someone who was the victim of a violent crime in their neighborhood, bringing the reality of violence very close to many of the Head Start families. Five percent of parents reported being a victim of violent crime in their neighborhood, and similarly, 5.0 percent of the parents reported being victims of violence in their homes.

Exposure to violence varied across ethnic groups. Among parents of African American children, 34.5 percent reported seeing nonviolent crimes in their neighborhoods, a figure that was over twice the rate reported by parents of White children (15.0 percent) and approximately one third more than parents of Hispanic children (24.2 percent). This pattern held for each type of exposure to violent crime. Over twice as many parents of African American children (36.3 percent) reported witnessing violent crimes in their neighborhoods compared to parents of White children (15.5 percent) and Hispanic children (17.5 percent). For reports of victimization, parents of African American children were again highest, with 7.4 percent indicating they were victims of crime in their neighborhoods, and 6.2 percent reporting they were victims of violent crime in their homes. These victimization rates were closer to reports for parents of Hispanic children (5.8 percent in the neighborhoods; 5.4 percent in their homes) than for parents of White children (4.4 percent in the neighborhoods; 4.5 percent in their homes).

As for the Head Start children, 3.8 percent were reported by parents to have witnessed a violent crime and 8.6 percent were reported to have witnessed domestic violence during the previous year. Slightly over one percent of the children were reported by their parents to have been victims of violent crime (1.1 percent), while almost two percent were victims of domestic violence (1.7 percent) during the previous year.

Exposure to violence had direct and indirect associations with child outcomes. Partial correlations controlling for parental income, education, employment, and child gender, age, and ethnicity revealed small but significant positive correlations between parents' reports of exposure to violence and parents' reports of child problem behavior. In these analyses, exposure to violence is represented as a summary score of how often each of the five types of exposure to violence (as noted above) was reported by parents. Scores ranged from 5 (no exposure) to 15 (more than one exposure to every type). The mean score for the sample was 6.1. Parents who reported greater exposure reported their children engaged in fewer positive social behaviors and more overall problem behavior; teachers of these children also reported they were more aggressive. This contrasts with the parents with less violence exposure, who reported more positive behaviors for their children. Exposure to violence did not have a direct relationship with child cognitive outcomes. Parents who reported more exposure to violence were significantly more depressed, but interestingly, they also were more likely to have an authoritative parenting style (e.g., less harsh, more use of rationales).

Controlling for mother's education and employment, household income, and child's race, age, and gender, multivariate logistic regression models were used to further explore the risks associated with exposure to violence. Odds ratios and 95% confidence intervals for risk factors (alcohol use, single parent, domestic violence, criminal justice system involvement) included in the models are presented in Table 6.2 (located at the end of the chapter). Parents who reported exposure to violence in their neighborhoods or homes, compared to parents who were not exposed to violence, were 1.5 times more likely to be single parents as well as to live in households where either they or someone else drank alcohol. Parents who were exposed to violence in their neighborhoods or homes, compared to parents not exposed to violence, were also 1.7 times more likely to have reported that they, another household member, or a non-household biological parent had been arrested or charged with a crime since the birth of their Head Start children, and over 2.7 times more likely to have screened positive for domestic violence (see measure below).

**Domestic Violence.** A three-item screening measure for domestic violence was administered to the parents (Feldhaus, Loziol-McLain & Amsury, 1997). They were asked if they had ever been hit, kicked, punched or otherwise hurt by anyone within the past year, if they felt safe in their current relationship, or if they currently felt unsafe from a partner in a previous relationship. Almost 13 percent of the parents answered 'yes' to one of these questions, thereby screening positively for experiencing domestic violence. Differences on family and child outcomes were found between families with a parent who screened positively for domestic violence versus those who did not. For example, parents experiencing domestic violence were significantly more depressed, reported their children to be more aggressive, more hyperactive, more withdrawn and to have more overall problem behavior. Teachers also reported these children to be more withdrawn and have more overall problem behavior than children whose parents were not experiencing domestic violence. However, no direct relationship was found between parent-reported domestic violence and child cognitive outcomes.

Multivariate logistic regression models, controlling for mother's education and employment, household income, and child's race, age, and gender, estimated that parents who screened positive for domestic violence, compared to those who did not, were 2.5 times more likely to be single parents, 1.5 times more likely to drink alcohol or live with a drinker, and almost 3.5 times more likely to report that they, another household member, or a non-household biological parent had been arrested or charged with a crime since the birth of their Head Start children..

**Substance Use in the Home.** The occurrence of substance use in homes, whether cigarette smoking, alcohol use, or drug use, is another challenge that faced a number of Head Start families. Almost one half of the Head Start children (45.1 percent) lived in households with at least one individual who smoked cigarettes. Smoking varied by ethnicity. Cigarette smoking was reported less frequently in households where African American children lived (38.8 percent) than in households of Hispanic children (52.2 percent) or White children (54.3 percent). More than one fourth of the parents (28.1 percent) reported drinking alcohol such as beer, wine, or liquor in the past 30 days; 20.0 percent drank less than once a week, and 7.8 percent reported drinking between 1-2 times per week to every day. Slightly over 40 percent of all households reported having at least one individual who drank alcohol. Among families who lived in households where someone drank alcohol, 7.9 percent reported alcohol-related problems with family members, 5.9 percent experienced trouble with the police because of alcohol, and 4.2 percent missed work or school due to alcohol-related illness. Less than one percent of the families reported having anyone in the household who used drugs.

The findings suggest that living in a household with someone who drinks increased the risk of negative family and child outcomes. Parents who drank or who lived with a drinker were significantly more likely to be depressed and to report their children to have more overall problem behavior than parents who did not drink or live in a household with a drinker. Children who lived with someone who drank alcohol scored lower on vocabulary, color naming, and social awareness tasks.

Further exploration of the risks associated with drinking or having a drinker in the household was done using multivariate logistic regression models, again controlling for mother's education and employment, household income, and child's race, age, and gender. Odds ratios and 95% confidence intervals for the associated risk factors included in the models are presented in Table 6.2. Parents who drank alcohol or lived with drinker were approximately 1.5 times more likely to report involvement in the criminal justice system, exposure to violence, or domestic violence in their lives compared to parents who did not drink or live with a drinker. Parents who drank or lived with a drinker were no more likely to be single parents than those who lived in alcohol-free households.

**Involvement With the Criminal Justice System.** In order to assess how many families had involvement with the criminal justice system, parents were asked if they, another household member, or a non-household biological parent had been arrested or charged with a crime since the birth of their Head Start children. Almost one fifth (19.2 percent) of the parents reported that someone had been arrested and charged with a crime and 16.7 percent reported someone who spent time in jail. Involvement with the criminal justice system did not vary much by ethnicity. About one fifth of parents of White children, parents of Hispanic children, and parents of African American children reported having someone in their family who was arrested (21.8 percent, 20.1 percent, and 18.9 percent, respectively).

T-tests were conducted to examine if child behavior and cognitive outcomes differed between children who were members of families with involvement in the criminal justice system and children from families without such involvement. Children from families who had someone arrested scored lower on assessed vocabulary and were reported by their parents to be more aggressive, more hyperactive, more withdrawn, and to have more overall problem behavior. They were also reported by their teachers to be more aggressive and have more overall problem behavior than children from families who did not have someone arrested.

Multivariate logistic regression models, controlling for mother's education and employment, household income, and child's race, age, and gender were used to determine estimates of risk among those families who had someone close to them who was involved in the criminal justice system (odds ratios and 95% confidence intervals are reported in Table 6.2). These risk estimates indicated that parents who reported that they, another household member, or a non-household biological parent had been arrested or charged with a crime since the birth of their Head Start children were twice as likely to be depressed, over 2.5 times more likely to be single parents, and over 3.5 times more likely to have been a victim of domestic violence than parents who did not have someone in their families involved in the criminal justice system.

Children in families from which someone had been arrested were at great risk for witnessing or being a victim of violent crime and domestic violence compared to children in families where no one had been arrested or charged with a crime. These Head Start children were more than 3 times more likely to have been a witness to violent crime and almost 3.5 times more likely to have witnessed domestic violence in the past year. Their risk of victimization also increased greatly. These children were 3 times more likely to have been both a victim of domestic violence and a victim of violent crime in the past year than children whose families did not report involvement with the criminal justice system. While there is a strong literature to suggest that children of incarcerated parents are at increased risk for negative outcomes (Lange, 2000), these findings suggest that risk greatly increases when there is any family involvement with the criminal justice system.

**Cumulative Social and Economic Risk.** Previous work has noted that social and economic risk factors may be associated with negative child outcomes, particularly when multiple risks are present (Annie E. Casey Foundation, 1999; ACYF, 2002). To explore this further, we adapted a standard list of family risk factors (Annie E. Casey Foundation, 1999) to investigate relationships with child outcomes. These risk factors include:

- The child lived with a single parent;
- The mother was a high school dropout;
- The family income was below the poverty line;
- The child was living with a parent(s) who did not have steady, full-time employment;
- The family was receiving welfare benefits; and
- The child did not have private health insurance.

While a number of these risks are already included in broader analytic models looking at the child outcomes (see earlier chapters), the cumulative effects of these risks may be associated with reduced school readiness. Across all six risks, at least one risk was evident in 89 percent of the Head Start families. For all families, the most prevalent risks were having a mother without a high school education (38.2 percent) being in a single-parent household (48.2 percent), and children not having private health insurance (64.6 percent).[23] None of the other risks were reported for more than 30 percent of the families.

Of greater concern is that increases in the number of risk factors, particularly counts of four or more risks, increased the likelihood of negative child outcomes. About one fifth of the families (20.3 percent) were found to have four or more risk factors, with almost 10 percent having five or six risks. Less than one fifth of the families of White children (18.6 percent) and Hispanic children (18.6 percent) had 4 or more of the risk factors, while one quarter of the families of African American children (24.9 percent) were found to have this high level of risk.

Parents reporting four or more risk factors also had higher depression scores, a lower locus of control, and, interestingly, scored higher on both the authoritarian and authoritative parenting style scales. These same parents also reported more problem behavior for their children, including aggressive and withdrawn behavior, and gave lower ratings on emergent literacy. In addition, the Head Start teachers also rated the children from families with four or more risks as having more withdrawn and hyperactive behavior.

---

[23] The Fall 2000 FACES parent interview identified children without private health insurance, but was not clear on whether children received Medicaid or SCHIP.

In terms of the child cognitive assessments, the cumulative risk of four or more risk factors was clearly associated with lower child outcomes. This high level of risk meant significantly poorer performance on social awareness, design copying, color naming, one-to-one counting, book knowledge, comprehension, print concepts, vocabulary, letter identification, early math, and early writing.

## B. Family Involvement and its Relationship to School Readiness

**Family Activities.** While the findings presented above suggest that a number of family characteristics are likely to have negative associations with children's school readiness, the frequency of families' interactions with their children reflects family strengths that could improve school readiness. For example, do families that are more active with their children have children with better behavior and better cognitive skills?

Parents were asked how often family members engaged in selected activities over the previous week as well as the previous month with their Head Start children. The weekly activities included telling the child a story; teaching letters, words, or numbers; teaching songs or music; doing arts and crafts; playing toys or games; doing errands; or doing household chores. Parents were also asked about how often they read to their children over the past week, and findings related to this variable are addressed in Chapter V. The monthly activities included visiting the library, shows, museums, and zoos; attending community or sporting events; and discussing family history. Total activity scores are based on the sum of the weekly and monthly activity scores.

The total activity score for combined weekly and monthly activities indicated that families engaged in a mean of 7.6 activities with the children, out of a possible 14 activities (National Center for Education Statistics, 1996). Weekly activities made up most of that total, with a reported mean of 5.7 activities of a possible seven, while a mean of 1.9 monthly activities was reported, also out of a possible seven.

Ethnic differences were noted in the number of activities families engaged in with their children. For the activities in the previous week, African American children had higher family activity than both White or Hispanic children, and White children had higher activity scores than Hispanic children. For the activities covered for the previous month, there was a similar effect for ethnicity, as African American children were engaged in more activity than White or Hispanic children.

There were several positive relationships between activities with children and child behavioral and cognitive outcomes. Partial correlations were run controlling for parent income, education, and employment; child gender, age, and ethnicity; and how often the child was read to in the previous week. The frequency of weekly, monthly, and combined family activities was positively correlated with parent reports on positive child behaviors and emergent literacy skills. All three types of activity scores were negatively correlated with parent reports of aggressive behavior, while weekly and total behavior scores were negatively associated with overall problem behavior and the hyperactive behavior subscale. Among parents, being engaged in activities with their children was correlated positively with both locus of control and authoritative parenting. Teacher ratings of hyperactive and overall problem behavior were negatively correlated with monthly activities.

With regard to child cognitive scores, partial correlations identified a number of positive relationships, with weekly activities having the strongest association with cognitive skills. For example, engaging in more activities with the children during the previous week was positively correlated with higher scores for the children on the color naming and vocabulary assessments. Participating in the monthly activities also had small positive correlations with the social awareness, color naming, one-to-one counting, book knowledge, and print concepts assessments.

**Parenting Style.** FACES 2000 provided the opportunity to look at the correlations of parenting style with selected child cognitive and behavioral outcomes. The parents were scored on two different parenting styles: authoritarian and authoritative. The first of these styles is generally considered a stricter, directive style, while the authoritative style reflects a less harsh style with greater use of rationales. The differences in these styles were apparent in their relationships with the parent and child outcomes. As expected, the authoritative and authoritarian scales were negatively correlated with each other.

Partial correlations, controlling for parent income, education, and employment; child gender, age, and ethnicity; and how often the child was read to in the previous week, were used to assess the relationship between parenting style and child behavioral and cognitive outcomes. Parents who scored higher on the authoritative subscale also had higher internal locus of control (i.e., felt more in control of their lives rather than at the mercy of external factors) and gave their children higher ratings of positive social behavior and emergent literacy. In contrast, the authoritarian subscale scores were positively correlated with the overall ratings of children's problem behavior, as well as the subscales of aggression, hyperactivity, and withdrawn behavior, but negatively correlated with parents' locus of control.

There were generally weak relationships between parenting style and the cognitive outcomes. The authoritative parenting style scores were positively correlated with children's social awareness and the authoritarian parenting style was negatively correlated with color naming, but neither parenting style was correlated with any of the other cognitive measures. Hyperactive behavior as rated by teachers was positively correlated with an authoritarian parenting style.

**Family Support From Head Start.** In the spring of 2001, parents were asked about the ways that they were involved in the Head Start program throughout the past school year. Parents most frequently reported attending parent-teacher conferences (79.0 percent), observing in their children's classrooms for at least 30 minutes (74.9 percent), and meeting with a Head Start staff member in their homes (69.4 percent). More than one half of the parents volunteered in their children's classrooms (60.0 percent) or prepared food or materials for special events (58.1 percent) and slightly less than one half helped with field trips (42.5 percent), or attended bazaars (42.3 percent) and workshops (42.5 percent). Less than one fourth of the parents participated in Policy Council (22.5 percent).

A summary score measuring total involvement was created for each parent who responded to the questionnaire. A series of partial correlations, controlling for parent income, education, and employment; child gender, age, and ethnicity; were conducted to examine the relationship between involvement at Head Start and other family and child factors. Higher levels of involvement at Head Start were significantly related to positive outcomes for the children. Parents who were more involved at Head Start were likely to report more positive social behavior for their children and less aggressive and overall problem behavior, as well as higher emergent literacy skills than parents who were less involved at Head Start. There was also a relationship between involvement and teachers' reports of the children's skills and with the children's scores on cognitive assessments. For example, teachers rated the children of more involved parents as more creative and more socially aware. On the cognitive measures, children whose parents were more involved at Head Start scored higher on vocabulary, book knowledge, early writing, early math, and letter-identification tasks.

Consistent with previous findings (ACYF, 2002), most parents reported high levels of satisfaction with the Head Start program. For example, while more than one half were very satisfied with Head Start in every area, more than 82 percent of the parents reported they were very satisfied with how the program helped their children grow and develop, respected the family's culture, identified child services, maintained a safe program, and prepared the children for kindergarten. There were very few unsatisfied parents, with reports of dissatisfaction generally under 2.5 percent. The area with the greatest degree of reported dissatisfaction was how Head Start helped parents become involved in their community, but even this concern was cited by less than 7 percent of the parents.

Parents also reported that they and their children had positive experiences at Head Start. Almost 85 percent of the parents reported that their children 'often' or 'always' received individual attention, while more than 90 percent reported their children 'often' or 'always' felt safe and secure at Head Start, were happy in the program, were treated with respect, and felt accepted by their teachers. With regard to the teachers, more than 90 percent of the parents reported that their children's teachers 'often' or 'always' were open to new learning, were warm towards the children, were interested in the children, were supportive to parents, were welcoming, did not use harsh discipline, and were happy.

### C. Head Start's Protective Role for Families and Children

Earlier findings from the FACES 1997 cohort of families suggest that Head Start may play an important role in protecting families and children from the challenges that many families face. To validate this finding in the FACES 2000 cohort, a series of linear regression models, controlling for parent income, education, and employment; child gender, age, and ethnicity; and parent level of activity with their children, tested whether the effect of maternal depression, exposure to violence, domestic violence, alcohol use, or involvement in the criminal justice system on child behavior and child cognitive outcomes varied as a function of (or was moderated by) Head Start satisfaction, experience, or involvement. Summary scores were created for parents' involvement at Head Start, for parents' reports of having a positive experience at Head Start, and parents' reports of satisfaction with the program. Means and standard deviations for these three variables are found in the Appendix.

Table 6.1 presents the interaction terms found to be significant moderators of risk factors in the regression equations. These findings support the earlier results and suggest that Head Start may play an important role in protecting families from the negative outcomes associated with these challenges. For example, there is a direct positive relationship between domestic violence and parent and teacher reports of children's problem behavior. This relationship remains consistent for children of parents who had fewer positive experiences with the program; however, when parents had positive experiences at Head Start, the effects of domestic violence on parent and teacher reports of increased aggressive, hyperactive, withdrawn, and overall problem behavior for children were no longer evident. These findings indicate that the negative effects of being exposed to domestic violence on children were buffered by parents' positive experiences at Head Start.

The moderating effects of Head Start on negative outcomes for children are consistent across all risk factors presented in Table 6.1, and include moderation of the relationship between risk factors and lower cognitive outcomes, as well as negative child behavior. For example, parents' depression, exposure to violence, or involvement in the criminal justice system were related, in some cases, to poorer cognitive outcomes for children, such as early math, book knowledge, color naming, one-to-one counting, and vocabulary. Having a parent who reported positive Head Start experiences or who was more involved with the program moderated this relationship and appeared to protect the children from these negative outcomes. All significant modifying relationships are presented below.

**Table 6.1: Significant Head Start Moderators of Negative Outcomes Associated With Risk Factors**

**Head Start Satisfaction, Experience, and Involvement as Moderators of the Negative Outcomes for Children and Families Related to Maternal Depression**

| Interaction Term (Independent Variable x Moderator) | Outcome Variable | Sig |
|---|---|---|
| Depression x Head Start Satisfaction | Withdrawn Behavior (parent report) | $p \leq .01$ |
| Depression x Head Start Involvement | Book Knowledge | $p \leq .05$ |
| Depression x Head Start Experience | Early Math | $p \leq .01$ |
| Depression x Head Start Experience | Creativity (teacher report) | $p \leq .05$ |

**Head Start Satisfaction, Experience, and Involvement as Moderators of the Negative Outcomes for Children and Families Related to Exposure to Violence**

| Interaction Term (Independent Variable x Moderator) | Outcome Variable | Sig |
|---|---|---|
| Violence x Head Start Involvement | Hyperactive Behavior (parent report) | $p \leq .05$ |
| Violence x Head Start Involvement | Emergent Literacy (parent report) | $p \leq .01$ |
| Violence x Head Start Involvement | Social Awareness | $p \leq .01$ |
| Violence x Head Start Involvement | Color Naming | $p \leq .01$ |
| Violence x Head Start Experience | One-to-One Counting | $p \leq .05$ |

**Head Start Satisfaction, Experience, and Involvement as Moderators of the Negative Outcomes for Children and Families Related to Domestic Violence**

| Interaction Term (Independent Variable x Moderator) | Outcome Variable | Sig |
|---|---|---|
| Domestic Viol x Head Start Experience | Social Skills (parent report) | $p \leq .05$ |
| Domestic Viol x Head Start Experience | Total Problem Behavior (parent report) | $p \leq .001$ |
| Domestic Viol x Head Start Experience | Aggressive Behavior (parent report) | $p \leq .05$ |
| Domestic Viol x Head Start Experience | Withdrawn Behavior (parent report) | $p \leq .01$ |
| Domestic Viol x Head Start Satisfaction | Hyperactive Behavior (parent report) | $p \leq .05$ |
| Domestic Viol x Head Start Experience | Total Problem Behavior (teacher report) | $p \leq .05$ |
| Domestic Viol x Head Start Experience | Social Score (teacher report) | $p \leq .05$ |

**Head Start Satisfaction, Experience, and Involvement as Moderators of the Negative Outcomes for Children and Families Related to Alcohol Use by Family Member**

| Interaction Term | Outcome Variable | Sig |
|---|---|---|

| (Independent Variable x Moderator) | | |
|---|---|---|
| Alcohol Use x Head Start Experience | Social Skills (parent report) | $p \leq .01$ |
| Alcohol Use x Head Start Satisfaction | Total Problem Behavior (parent report) | $p \leq .05$ |
| Alcohol Use x Head Start Experience | Withdrawn Behavior (parent report) | $p \leq .01$ |
| Alcohol Use x Head Start Satisfaction | Hyperactive Behavior (parent report) | $p \leq .01$ |
| Alcohol Use x Head Start Experience | Total Problem Behavior (teacher report) | $p \leq .01$ |
| Alcohol Use x Head Start Experience | Aggressive Behavior (teacher report) | $p \leq .01$ |
| Alcohol Use x Head Start Involvement | Emergent Literacy (parent report) | $p \leq .05$ |
| Alcohol Use x Head Start Experience | Social Score (teacher report) | $p \leq .05$ |

**Head Start Satisfaction, Experience, and Involvement as Moderators of the Negative Outcomes for Children and Families Related to Involvement with the Criminal Justice System**

| Interaction Term (Independent Variable x Moderator) | Outcome Variable | Sig |
|---|---|---|
| Crime x Head Start Experience | Social Skills (parent report) | $p \leq .05$ |
| Crime x Head Start Experience | Withdrawn Behavior (parent report) | $p \leq .01$ |
| Crime x Head Start Involvement | Vocabulary | $p \leq .05$ |

The difficult challenges or risks that parents face--depression, exposure to violence, domestic violence, alcohol use, and involvement with the criminal justice system--are often associated with negative behavioral and cognitive outcomes for their children. The findings cited above provide some evidence that Head Start may play an important role in protecting children and their families from the consequences of these risk factors. Parent involvement at Head Start, parent reports that they and their children had positive experiences at Head Start, and parent satisfaction with the program, significantly moderated relationships between these risk factors and many negative child behavior and cognitive outcomes. These findings provide support for the theory that children's school readiness is enhanced when programs work with families as well as with children.

## SUMMARY

This chapter examined family and parent characteristics that are related to school readiness in early childhood, and parents' important role in preparing children for school. The challenges families face and the strengths they possess are important considerations in understanding how best to prepare children for school. The main findings are summarized below.

- One quarter of the parents were classified as moderately or severely depressed. Parents who were more depressed reported that their children had more problem behaviors and fewer positive social behaviors, a finding supported by the teachers' reports of children's behavior. Their children also had lower scores on one-to-one counting, creativity, design copying, early writing, letter identification, and early math assessments.

- More than one fifth of the parents reported they had witnessed violent crime. Five percent reported being victims of violent crime in the neighborhood, while a similar percentage reported being victims of violence in their homes. Almost 10 percent of the children were reported to have witnessed domestic violence during the previous year. Less than two percent of the children were reported to have been victims of violent crime or victims of domestic violence. Positive correlations were found between increased exposure to neighborhood violence and reports of child problem behavior, while children in more violent neighborhoods had lower assessment scores on the color naming and book knowledge assessments.

- Almost 13 percent of the parents indicated that they have been victims of domestic violence. Teachers and parents reported children in these families had more overall problem behaviors. However, no direct relationship was found between parent reports of domestic violence and child cognitive outcomes.

- Almost one half of the Head Start children lived in households with at least one individual who smoked cigarettes and about two fifths of the households reported having at least one individual who drank alcohol. Living in a household with someone who drinks increased the risk of maternal depression, while the children in these homes were reported to have more overall problem behavior and scored lower on vocabulary, color naming, and social awareness assessments.

- Almost one fifth of the parents reported that someone in their household had been arrested and charged with a crime. Children in these families were more than three times more likely to have been a witness to violent crime or domestic violence in the past year. These children were also three times more likely to have been a victim of domestic violence or violent crime. These children had lower vocabulary scores, and were reported by both parents and teachers to be more aggressive and have more overall problem behaviors.

- At least one of a set of selected risk factors was evident in over 90 percent of the families. Almost one quarter of the families had four or more risk factors. Children in these families had lower parent ratings on emergent literacy and higher teacher and parent ratings of problem behavior. In the assessments, these children scored lower on design copying, color naming, one-to-one counting, book knowledge, vocabulary, early math, early writing, and letter identification.

- Families engaged their children in a number of weekly and monthly activities. The number of activities was positively correlated with positive child behaviors and emergent literacy and negatively correlated with problem behaviors. In particular, the weekly activities had positive correlations with scores on the social awareness, color naming, one-to-one counting, book knowledge, vocabulary, early math, early writing, and letter identification tasks.

- Higher authoritative parenting style scores were significantly positively correlated with children's social awareness, but not with any of the other cognitive measures. On the other hand, higher authoritarian scores were significantly negatively correlated with comprehension, color naming, vocabulary, and early math assessments.

- More than two thirds of parents had attended parent-teacher conferences, observed in their children's classrooms for at least 30 minutes, or met with a Head Start staff member in their homes. Parent involvement in Head Start was positively correlated with parental reports of positive social behavior and higher emergent literacy skills and negatively correlated with aggressive and overall problem behavior. Children with more involved parents scored higher on vocabulary, book knowledge, early writing, early math, and letter identification tasks.

- Head Start may play an important role in protecting families and children from the challenges that low-income families face. Parent involvement at Head Start, parent reports that they and their children had positive experiences at Head Start, or parent satisfaction with the program, significantly moderated relationships between risk factors such as maternal depression, exposure to violence and domestic violence, substance use, and involvement with the criminal justice system, and many negative child behavior and lower cognitive outcomes.

## D. REFERENCES

Annie E. Casey Foundation (1999). *1999 Kids count data book: State profiles of child well-being.* Baltimore, MD: Annie E. Casey Foundation.

Administration on Children, Youth, and Families (2002). *A Descriptive Study of Head Start Families: FACES Technical Report I.* Washington, D.C.: U.S. Department of Health and Human Services.

Feldhaus, K.M., Koziol-McLain, J., Amsbury, H.L. (1997). Accuracy of 3 brief screening questions for detecting partner violence in the emergency department. *Journal of the American Medical Association, 227(17)*, 1357.

Hall, L., Williams, C., & Greenberg, R. (1985). Supports, stressors, and depressive symptoms in low-income mothers of young children. *American Journal of Public Health, 75*, 518-522.

Lange, S.M. (2000). The challenges confronting children of incarcerated parents. *Journal of Family Psychotherapy, 11(4)*, 61-68.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126(2),* 309-337.

Liaw, F., & Brooks-Gunn, J. (1994). Cumulative familial risks and low birth weight children's cognitive and behavioral development. *Journal of Clinical Child Psychology, 23*, 360-372.

National Center for Education Statistics. (1996). National Household Education Survey Parent and Family Involvement Questionnaire. Washington, DC: U.S. Department of Education.

Radloff, L.S. (1977). The CES-D scale: A self report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385-401.

National Center for Education Statistics. (1996). National Household Education Survey Parent and Family Involvement Questionnaire. Washington, DC: U.S. Department of Education.

**Table 6.2 Significant Adjusted Odds Ratios and 95% Confidence Intervals for Maternal Depression, Exposure to Violence, Domestic Violence, Alcohol Use, and Involvement With the Criminal Justice System**

| | Maternal Depression | |
|---|---|---|
| Risk Factors | Adjusted Odds Ratio | 95% Confidence Interval |
| | | |
| Exposure to Violence | 1.69 | (1.38 – 2.06) |
| Domestic Violence | 2.99 | (2.33 – 3.84) |
| Alcohol Use | 1.36 | (1.12 – 1.65) |
| Criminal Justice System Involvement | 1.97 | (1.57 – 2.47) |
| Single Parent | 1.45 | (1.17 – 1.79) |
| | | |
| | Exposure to Violence | |
| Risk Factors | Adjusted Odds Ratio | 95% Confidence Interval |
| | | |
| Domestic Violence | 2.68 | (2.09 – 3.44) |
| Criminal Justice System Involvement | 1.73 | (1.38 – 2.16) |
| Alcohol Use | 1.46 | (1.21 – 1.75) |
| Single Parent | 1.43 | (1.18 – 1.74) |
| | | |
| | Domestic Violence | |
| Risk Factors | Adjusted Odds Ratio | 95% Confidence Interval |
| | | |
| Criminal Justice System Involvement | 3.47 | (2.68 – 4.50) |
| Single Parent | 2.61 | (1.97 – 3.46) |
| Alcohol Use | 1.50 | (1.18 – 1.91) |
| | | |
| | Alcohol Use | |
| Risk Factors | Adjusted Odds Ratio | 95% Confidence Interval |
| | | |
| Criminal Justice System Involvement | 1.48 | (1.19 – 1.83) |
| Exposure to Violence | 1.45 | (1.21 – 1.74) |
| Domestic Violence | 1.48 | (1.17 – 1.89) |
| | | |
| | Criminal Justice System Involvement | |
| Risk Factors | Adjusted Odds Ratio | 95% Confidence Interval |
| | | |
| Maternal Depression | 1.97 | (1.57 – 2.47) |
| Single Parent | 2.64 | (2.06 – 3.39) |
| Domestic Violence | 3.47 | (2.68 – 4.50) |
| Child Witness to Violence | 3.23 | (2.04 – 5.12) |
| Child Witness to Domestic Violence | 3.60 | (2.63 – 2.90) |
| Child Victim of Violence | 3.06 | (1.35 – 6.97) |
| Child Victim of Domestic Violence | 3.14 | (1.57 – 6.30) |

# Chapter VII: Predictive Validity of Cognitive and Behavioral Measures: Relationships Within and Across Cognitive and Social Developmental Domains

The child assessment battery for the Family and Child Experiences Survey (FACES) was designed to provide a comprehensive description of Head Start children's development in both the cognitive and socio-emotional domains. As an indicator of the validity of the FACES instruments, criterion-related validity, that is the degree to which the test or questionnaire correlates with one or more outcome criteria, was assessed. Systematic evidence for criterion validity is often described in terms of *predictive* validity. Predictive validity is assessed when the outcome criterion is measured at a later time point after the evaluated test (e.g., later school year). This chapter evaluates how well the FACES measures predict children's achievement and school adjustment at the end of their kindergarten year.[24]

## RESEARCH QUESTIONS

This chapter will address the following research questions:

1. How well does the FACES battery predict children's early reading skills and general knowledge at the end of kindergarten?
2. How well do the FACES parent and teacher ratings of children's social competence predict to children's school adjustment at the end of kindergarten?
3. Do the FACES parent and teacher ratings of children social competence contribute to the FACES battery's prediction of children's reading skills and general knowledge at the end of kindergarten?
4. How well do cognitive gains during Head Start predict reading and general knowledge at the end of kindergarten?
5. How well do changes in behavior ratings during Head Start predict social competence at the end of kindergarten?
6. Do FACES cognitive and behavior measures predict promotion to first grade?

## FINDINGS

### A. Children's Scores on the FACES Instruments at the End of Head Start Predict Kindergarten Outcomes

The FACES cognitive measures were designed to assess preliteracy skills, as well as general school readiness. In order to determine the predictive validity of these measures, relationships between the FACES assessment scale scores obtained during the Head Start year and the Reading and General Knowledge scale scores obtained at the end of kindergarten were examined by two approaches. In the first approach, Reading and General

---

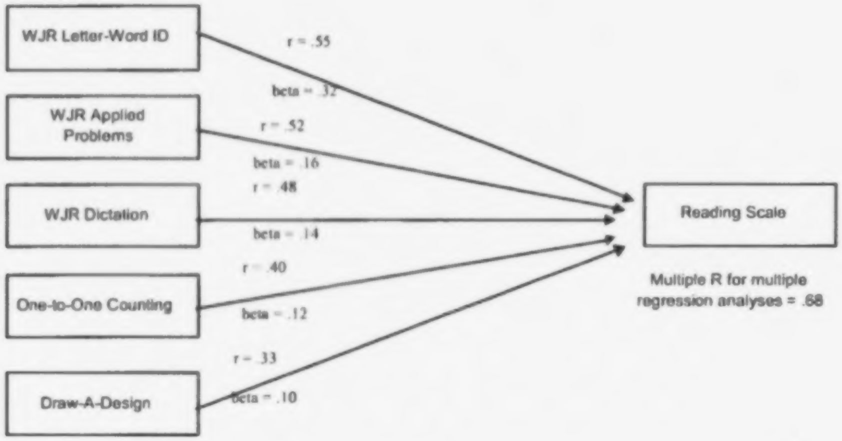
[24] Final kindergarten data from FACES 2000 are not yet available, therefore the following analyses are conducted with FACES 1997-1998 data. All of the described analyses were conducted on children who were assessed in English at all timepoints.

Knowledge scale scores from the end of the kindergarten year were correlated with the scores from each of the FACES subtests from the end of the Head Start year. The second approach assessed the ability of the FACES scale scores from the end of the Head Start year to predict Reading and General Knowledge scale scores at the end of the kindergarten year in two multiple regression analyses.

Children's scores on each of the component tasks in the FACES battery at the end of Head Start correlated significantly with their Reading scale scores at the end of kindergarten. Bivariate correlations with the Reading scale ranged from .55 (for the Woodcock-Johnson-R Letter-Word Identification) to .23 (for the Social Awareness). These moderate to high correlations indicate that the FACES measures have predictive power on outcome criteria at later time points. When the subtest scores were combined in a multiple regression model, the model did quite well at predicting children's early reading skills at the end of kindergarten, accounting for 46 percent of the variance in Reading scale scores. The best predictor of Reading scale scores was the Woodcock-Johnson-R Letter-Word Identification task, which also showed the highest bivariate correlation with the Reading scale (Figure 7.1).

Similarly, children's scores on each of the component tasks in the FACES battery at the end of Head Start also correlated significantly with their scores on the General Knowledge scale at the end of kindergarten. Bivariate correlations with the General Knowledge scale ranged from .77 (for the Peabody Picture Vocabulary Test-III or PPVT-III) to .30 (for the Draw-a-Design and Social Awareness), again indicating that the FACES measures have predictive power on outcome criteria at later time points. When the sub-test scores were combined in a multiple regression model, the model accounted for 65 percent of

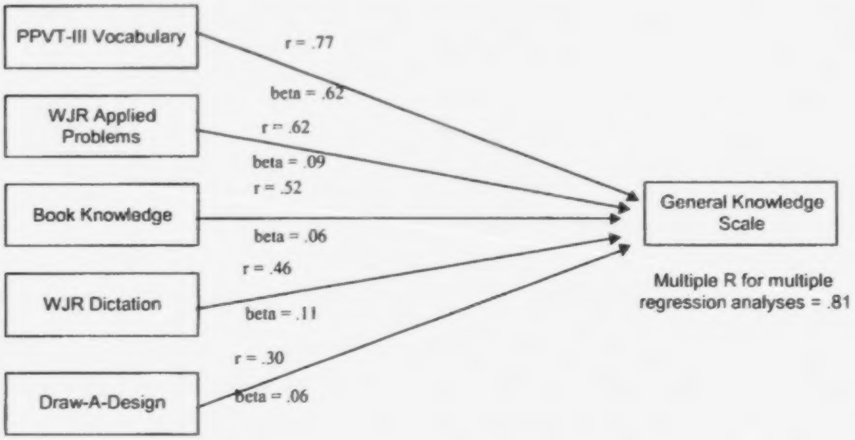
Figure 7.1. Correlations and Standardized Multiple Regression Coefficients Between Reading Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start Year



SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 1999 Kindergarten Followup, Ns = 718-786 for bivariate correlations, 623 for multiple regression coefficients.

the variance in General Knowledge scale scores. The best predictor in the multiple regression was the PPVT-III (beta = .62), which also showed the highest bivariate correlation with the General Knowledge scale (Figure 7.2).

Figure 7.2. Correlations and Standardized Multiple Regression Coefficients Between General Knowledge Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start Year



SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 199
Kindergarten Followup, Ns = 720-787 for bivariate correlations, 624 for multiple regression coefficients.

Neither the PPVT-III nor the Book Knowledge score were significant predictors of the Reading scale score in its multiple regression model, though scores on both of these tasks correlated significantly with the Reading score (r's = .42 and .39, respectively). It is noteworthy to also mention that the best predictor of the Reading score, namely Woodcock-Johnson-R Letter-Word Identification, was not a significant predictor of General Knowledge in its multiple regression analysis, despite the fact that the Letter-Word Identification task did correlate significantly with the General Knowledge task (r = .40). Conversely, the best predictor of General Knowledge, namely, the PPVT-III, did not have a significant role in the multivariate prediction of children's Reading scores.

It appears that the Reading and General Knowledge assessments may be tapping two distinct clusters of skills, both of which have been shown to be important for children's future reading proficiency and academic achievement. The skills tapped by the Reading scale during kindergarten were primarily what Whitehurst and Lonigan (1998) have called "inside-out" skills such as letter recognition, letter-sound association, and word decoding. The skills tapped by the General Knowledge assessment were "outside-in" skills, such as general information, word knowledge, and conceptual understanding that children need to help them comprehend and evaluate what they read and relate it to facts and concepts they have previously acquired. The FACES battery showed its validity by predicting well to children's later learning in both skill domains. In addition, these results suggest that development in both skills domains should receive attention in preschool curricula and practice, in order to foster both types of school-age abilities.

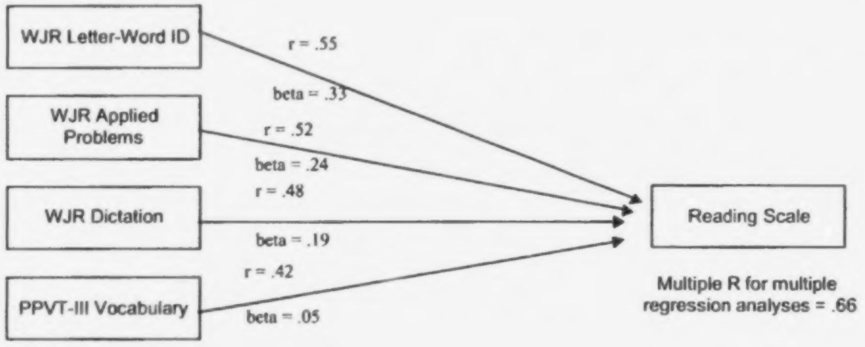
Although both domains were predicted well, the combination of subtests that produced the best forecasts differed across the two skill clusters. The Letter-Word Identification test was the best predictor of inside-out skills, with Applied Problems, Dictation, One-to-One Counting, and McCarthy Draw-a-Design contributing additional predictive power. The PPVT-III was by far the best predictor of outside-in skills, with Applied Problems, Book Knowledge, Dictation, and the McCarthy Draw-a-Design showing much smaller but significant regression coefficients as well. It is noteworthy that seven of the eight FACES subtests contributed significantly to either the Reading or General Knowledge regression model or both.

## Predictive Validity of Abbreviated FACES Battery

In order to further explore the validity of the FACES battery, multivariate regression analyses were carried out with the set of norm-referenced tests (i.e., PPVT-III and the Woodcock-Johnson-R subtests) at the end of Head Start predicting the Reading and General Knowledge scale at the end of kindergarten. These analyses were repeated with the set of criterion-referenced measures (i.e., Social Awareness, McCarthy Draw-a-Design, Color Names, One-to-One Counting, and Book Knowledge) as the predictors.

The four norm-referenced tests at the end of Head Start did almost as well as the full battery at predicting the Reading scores at the end of kindergarten (Figure 7.3), predicting 43 percent of the variance. The Letter-Word Identification was most closely associated with the Reading scores. Applied Problems and Dictation had significant coefficients as well. But, as before, the PPVT-III did not.

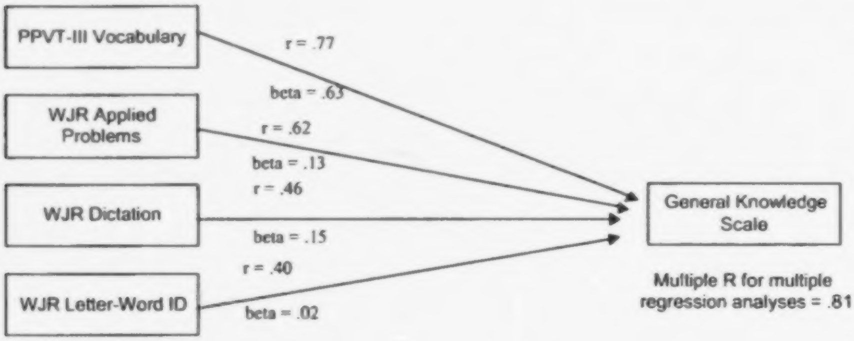
Figure 7.3. Correlations and Standardized Multiple Regression Coefficients between Reading Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start Year – Norm-Referenced Tests Only



SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 199 Kindergarten Followup, Ns = 718-737 forbivariate correlations, 640 for multiple regression coefficients.

The four norm-referenced tests at the end of Head Start also did as well as the full battery at predicting the General Knowledge scores at the end of kindergarten, predicting 65 percent of the variance. The PPVT-III was most closely associated with the General Knowledge scores. Dictation and Applied Problems had significant regression coefficients, but Letter-Word Identification did not (Figure 7.4).

Figure 7.4. Correlations and Standardized Multiple Regression Coefficients Between General Knowledge Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start – Norm-Referenced Tests Only

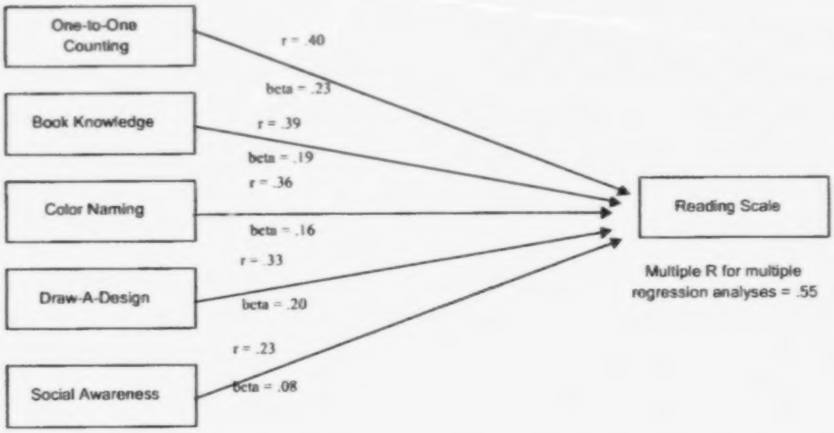


SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 199 Kindergarten Followup, Ns = 720-748 for bivariate correlations, 642 for multiple regression coefficients.

The five criterion-referenced FACES tasks significantly predicted the Reading and General Knowledge assessments at the end of kindergarten, though notably less well than either the full battery or the abbreviated battery composed of norm-referenced tests, predicting 30 percent and 36 percent of the respective variances. All five subtests were significant predictors of the Reading score (Figure 7.5). As with the full battery, tasks that tapped perceptual-motor skills and knowledge of print conventions were more closely related to kindergarten Reading scores than were tasks that tapped vocabulary or general information.

All five sub-tests were significant predictors of General Knowledge as well, but the pattern of relative magnitudes was different than that for Reading. Book Knowledge was most closely related with General Knowledge, while the other four sub-tests had relatively lower, but significant regression coefficients ranging from .15 to .13 (Figure 7.6).

Figure 7.5. Correlations and Standardized Multiple Regression Coefficients Between Reading Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start Year - Criterion-Referenced Tests Only

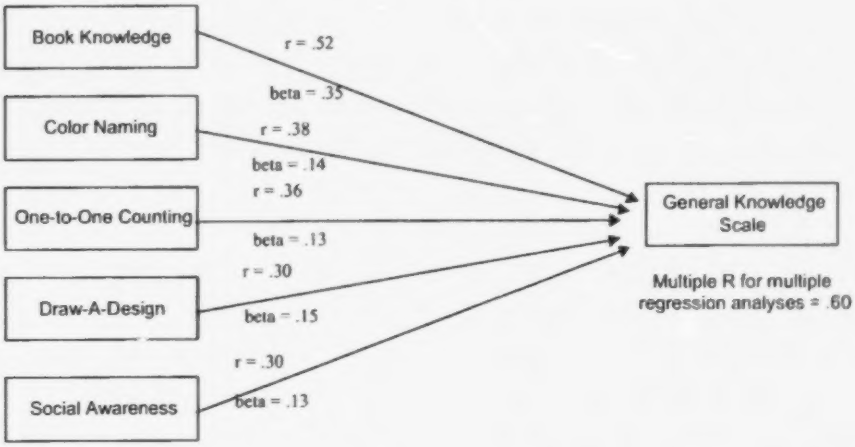

One-to-One Counting
r = .40
beta = .23
Book Knowledge
r = .39
beta = .19
Color Naming
r = .36
beta = .16
Draw-A-Design
r = .33
beta = .20
Social Awareness
r = .23
beta = .08
Reading Scale

Multiple R for multiple regression analyses = .55

SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 1999 Kindergarten Followup , Ns = 772-786 for bivariate correlations, 761 for multiple regression coefficients.

The combination of the four norm-referenced tests at the end of Head Start that were more extended (and hence more reliable in terms of internal consistency) did a better job of predicting to children's kindergarten achievement than the combination of the five FACES

Figure 7.6. Correlations and Standardized Multiple Regression Coefficients Between General Knowledge Scale Scores at End of Kindergarten Year and FACES Assessment Scale Scores at End of Head Start - Criterion-Referenced Tests Only


Book Knowledge
r = .52
beta = .35
Color Naming
r = .38
beta = .14
One-to-One Counting
r = .36
beta = .13
Draw-A-Design
r = .30
beta = .15
Social Awareness
r = .30
beta = .13
General Knowledge Scale

Multiple R for multiple regression analyses = .60

SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 199
Kindergarten Followup, Ns = 773-787 forbivariate correlations, 762 for multiple regression coefficients.

criterion-referenced tasks. However, further analyses revealed that the five criterion-referenced measures provide significant unique contributions to the prediction of the Reading scale score, over and above the contributions of the norm-referenced tests, increasing the predictive power of the assessment battery by almost 3 percent. The criterion-referenced measures also provided unique contributions to the prediction of the General Knowledge scale score increasing the predictive power by .6 percent, an association at the trend level ($p < .10$). These results indicate that the criterion-referenced measures significantly contribute to the assessment battery, by picking up variance not accounted for by the norm-referenced tests.

### B. FACES Behavioral Ratings at the End of Head Start Predict Children's Social Competence in Kindergarten

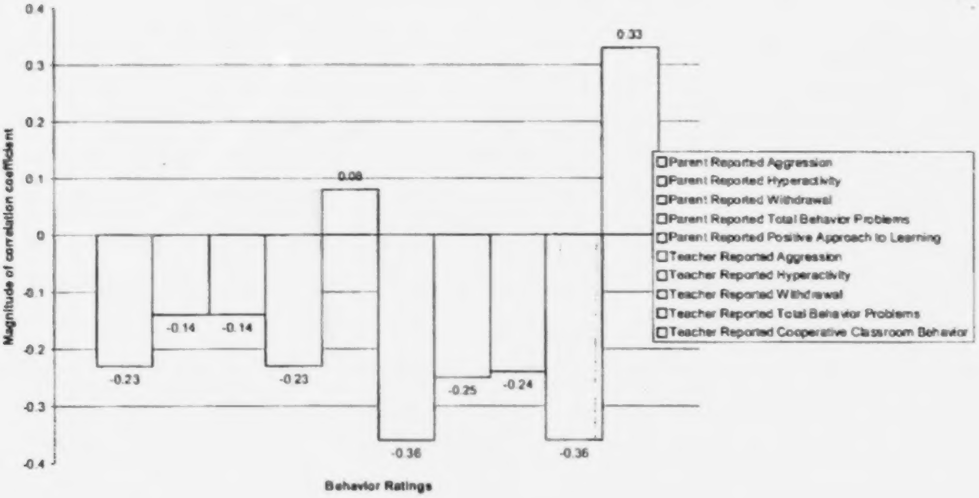
Social competence is an important developmental domain measured by the FACES battery. Two sources of information are tapped for assessing children's social competencies by collecting ratings of the children's behavior from their teachers and parents. These analyses examine the ability of the teacher and parent ratings of children's social competencies during Head Start to predict children's school adjustment at the end of kindergarten. Analyses of the predictive validity of the behavior ratings mirror those for the cognitive measures. First, teacher ratings of cooperative classroom behavior and total problem behaviors from the end of the kindergarten year were correlated with the behavior ratings from teachers and parents at the end of the Head Start year. Then the teacher and parent ratings were combined in multiple regressions predicting teacher ratings of cooperative classroom behavior and total problem behaviors at the end of the kindergarten year in two multiple regression analyses.[25]

Parent and teacher ratings of behavior at the end of Head Start were moderately correlated with teacher ratings of cooperative classroom behavior at the end of kindergarten, indicating that the FACES measures have predictive power on kindergarten outcomes (Figure 7.7). Correlations were significant in the expected directions. Problem behaviors as rated by both parents and teachers at the end of Head Start had significantly negative correlations with teacher ratings at the end of kindergarten, ranging from -.36 (for teacher reported ratings of total behavior problems and aggressive behavior) to -.14 (for parent reported ratings of hyperactive behavior and withdrawal behavior). Parent ratings of positive approaches to learning and teacher ratings of cooperative classroom behavior were both positively correlated with teacher ratings of cooperative classroom behavior at the end of kindergarten ($rs$ = .08 and .33 respectively). In general, teacher ratings showed stronger relationships than the parent ratings did with the kindergarten outcomes.

---

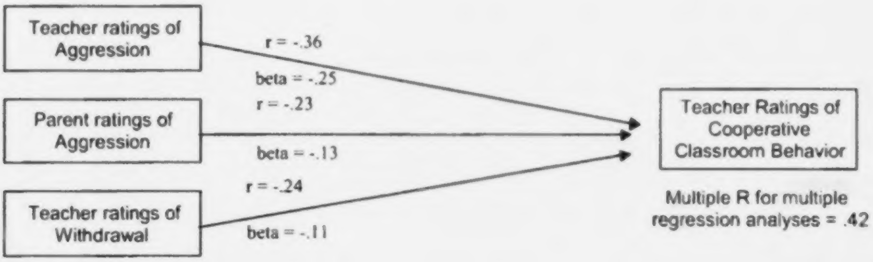
[25] In all multiple regression analyses involving the behavior ratings, teacher and parent ratings of total behavior were excluded as predictor variables from the models. Because these ratings are summative scores of the ratings for aggression, withdrawal, and hyperactivity, including them would introduce multicollinearity among these predictor variables to the model. Therefore, they were excluded from the regressions. They are included in the bivariate correlation analyses.

Figure 7.7. Correlation Coefficients Between Parent and Teacher Ratings of Children's Social Competence at the End of Head Start and Teacher Reported Ratings of Cooperative Classroom Behavior at the End of Kindergarten



When the ratings were combined in a multiple regression, the model accounted for 18 percent of the variance in teacher ratings of cooperative classroom behavior. The best predictor in the multiple regression was teacher reported ratings of aggressive behavior (beta = -.25), which also had the strongest bivariate correlation (Figure 7.8).

Figure 7.8. Correlations and Significant Standardized Multiple Regression Coefficients Between Teacher Ratings of Cooperative Classroom Behavior at End of Kindergarten Year and Parent and Teacher Behavior Ratings at End of Head Start Year



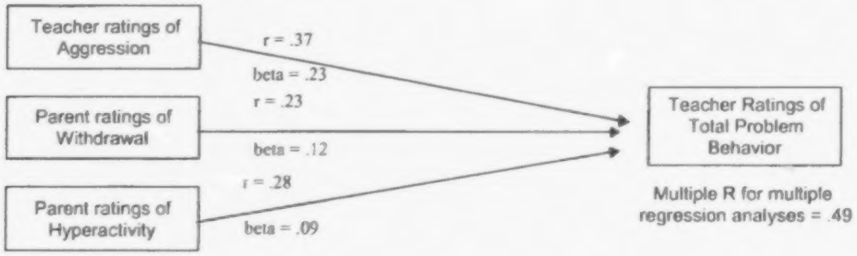
Teacher ratings of Aggression
r = -.36
beta = -.25
r = -.23
Parent ratings of Aggression
beta = -.13
r = -.24
Teacher ratings of Withdrawal
beta = -.11
Teacher Ratings of Cooperative Classroom Behavior
Multiple R for multiple regression analyses = .42

SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 199( Kindergarten Followup. Ns = 556-575 forbivariate correlations, 542 for multiple regression coefficients.

Similarly, parent and teacher ratings of behavior at the end of Head Start were moderately correlated with teacher ratings of total problem behaviors at the end of kindergarten, again indicating that the FACES measures have predictive power on kindergarten outcomes. Correlations were significant in the expected directions. Problem behaviors as rated by both parents and teachers at the end of Head Start had significantly positive correlations with teacher ratings of total problem behaviors at the end of kindergarten, ranging from .37 (for teacher reported ratings of aggressive behavior) to .23 (for parent reported ratings of aggressive and withdrawal behavior). Parent ratings of positive approaches to learning and teacher ratings of cooperative classroom behavior were both negatively correlated with teacher ratings of total problem behavior at the end of kindergarten (rs = -.15 and -.36 respectively). In general, teacher ratings showed stronger relationships than the parent ratings did with the kindergarten outcomes.

When the ratings were combined in a multiple regression, the model accounted for 24 percent of the variance in teacher ratings of total problem behavior. The best predictor in the multiple regression was teacher reported ratings of aggressive behavior (beta = .23), which also had the strongest bivariate correlation (Figure 7.9).

Figure 7.9. Correlations and Significant Standardized Multiple Regression Coefficients Between Teacher Ratings of Total Problem Behavior at End of Kindergarten Year and Parent and Teacher Behavior Ratings at End of Head Start Year


Teacher ratings of Aggression
r = .37
beta = .23
r = .23
Parent ratings of Withdrawal
beta = .12
r = .28
Parent ratings of Hyperactivity
beta = .09
Teacher Ratings of Total Problem Behavior
Multiple R for multiple regression analyses = .49

SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 199 Kindergarten Followup, Ns = 556-575 for bivariate correlations, 542 for multiple regression coefficients.

### Predictive Validity of Abbreviated Sets of Behavior Ratings

In order to further explore the validity of the behavior ratings, multivariate regression analyses were carried out with the set of teacher reported ratings at the end of Head Start predicting the teacher reported ratings of cooperative classroom behavior and total problem behavior at the end of kindergarten.

The teacher reported ratings at the end of Head Start did almost as well as the full set of behavior ratings at predicting the teacher reported ratings of cooperative classroom behavior at the end of kindergarten. The model explained 16 percent of the variance in kindergarten cooperative classroom behavior ratings, and the rating of aggressive behavior was most closely associated with the cooperative classroom behavior ratings (beta = -.26). The rating of cooperative classroom behavior had a significant regression coefficient as well (beta = .14). But, as before, the ratings of hyperactive and withdrawal behavior did not.

The four parent reported behavior ratings significantly predicted the teacher ratings of cooperative classroom behavior at the end of kindergarten, though notably less well than either the full set of behavior ratings or the set of teacher reported ratings. The model explained 6 percent of the variance in teacher reported cooperative classroom behavior and only parent reported ratings of aggressive behavior had a significant regression coefficient (beta = -.19).

The combination of the four teacher reported behavior ratings at the end of Head Start did a better job of predicting children's kindergarten behavior than the combination of the four parent reported behavior ratings. However, the four parent ratings provided significant unique contributions to the prediction of the cooperative classroom behavior ratings, over and above the contributions of the teacher ratings, increasing the predictive power of the assessment battery by almost 2 percent. These results indicate that the parent ratings significantly contribute to the assessment battery, by picking up variance not accounted for by the teacher reported ratings.

Similarly, the teacher reported ratings at the end of Head Start did almost as well as the full set of behavior ratings at predicting the teacher reported ratings of total problem behavior at the end of kindergarten. The model explained 18 percent of the variance in kindergarten total problem behavior ratings, and the rating of aggressive behavior was most closely associated with the total problem behavior ratings (beta = .22). The rating of cooperative classroom behavior had a significant regression coefficient as well (beta = -.17). But the ratings of hyperactive and withdrawal behavior did not.

The four parent reported behavior ratings also significantly predicted the teacher ratings of total problem behavior at the end of kindergarten, though notably less well than either the full set of behavior ratings or the set of teacher reported ratings. The model explained 11 percent of the variance in teacher reported total problem behavior and parent reported ratings of hyperactive behavior had the largest regression coefficient (beta = .18).

The combination of the four teacher reported behavior ratings at the end of Head Start did a better job of predicting children's kindergarten behavior than the combination of the four parent reported behavior ratings. However, the four parent ratings provided significant unique contributions to the prediction of the cooperative classroom behavior ratings, over and above the contributions of the teacher ratings, increasing the predictive power of the assessment battery by almost 6 percent. These results indicate that the parent ratings significantly contribute to the assessment battery, by picking up variance not accounted for by the teacher reported ratings.

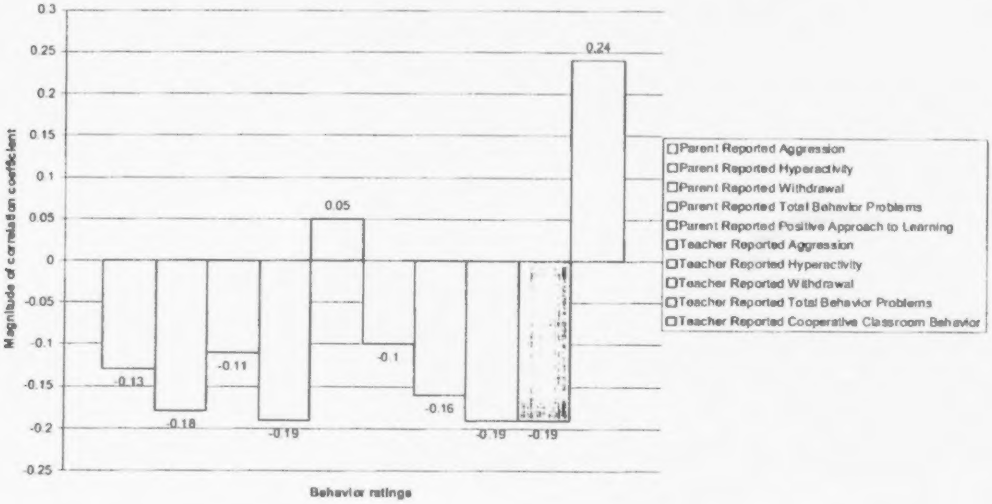
### C. Behavior Ratings at the End of Head Start Predict Reading Skills and General Knowledge at the End of Kindergarten

Given that certain positive behaviors may foster learning, while other negative behaviors may impede learning, the ability of these behavior ratings obtained during the Head Start year to predict Reading and General Knowledge scale scores obtained at the end of kindergarten was also examined.

Almost all of the parent and teacher ratings of behavior at the end of Head Start correlated significantly with Reading scale scores at the end of kindergarten, indicating that the

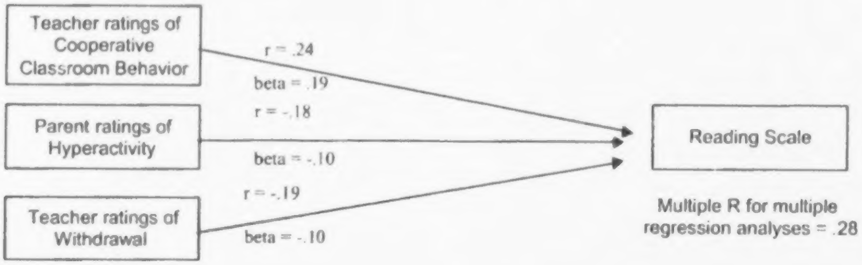
FACES measures have predictive power on outcome criteria at later time points. Only parent reported ratings of positive approaches to learning was not significantly related with Reading scores at the end of kindergarten. Correlations were significant in the expected directions. Ratings of problem behaviors were negatively correlated with kindergarten Reading scale scores, suggesting that behaviors that may impede learning are associated with lower reading skills in kindergarten (Figure 7.10).

Figure 7.10. Correlation Coefficients Between Parent and Teacher Ratings of Children's Social Competence at the End of Head Start and Reading Scale Scores at End of Kindergarten



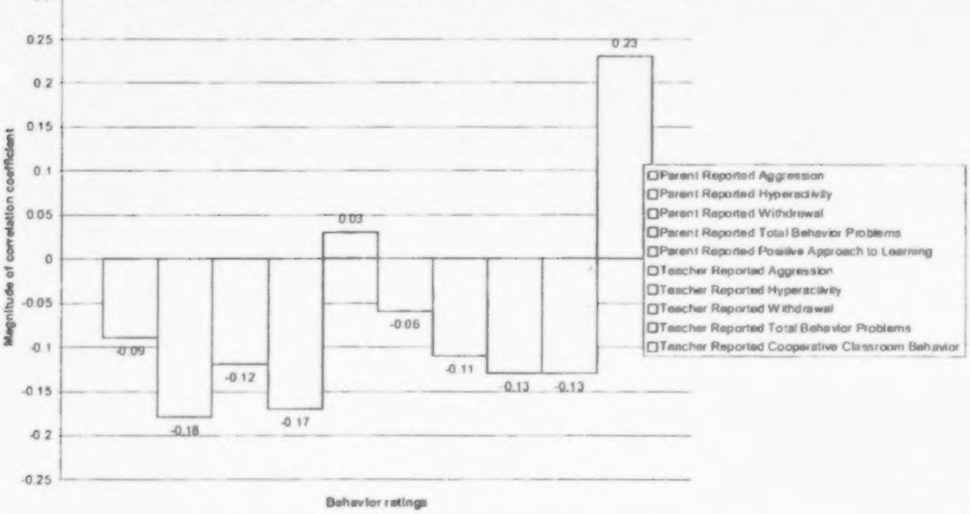
When the behavior ratings were combined in a multiple regression model, the model accounted for 8 percent of the variance in Reading scale scores. The best predictor in the multiple regression was teacher reported ratings of cooperative classroom behavior (beta = .19), followed by parent reported ratings of hyperactive behavior (beta = -.10), followed by teacher reported ratings of withdrawal behavior (beta = -.10) (Figure 7.11).

Figure 7.11. Correlations and Significant Standardized Multiple Regression Coefficients between Reading Scale Scores at End of Kindergarten Year and Parent and Teacher Behavior Ratings at End of Head Start


Teacher ratings of Cooperative Classroom Behavior
r = .24
beta = .19
Parent ratings of Hyperactivity
r = -.18
beta = -.10
Teacher ratings of Withdrawal
r = -.19
beta = -.10
Reading Scale
Multiple R for multiple regression analyses = .28

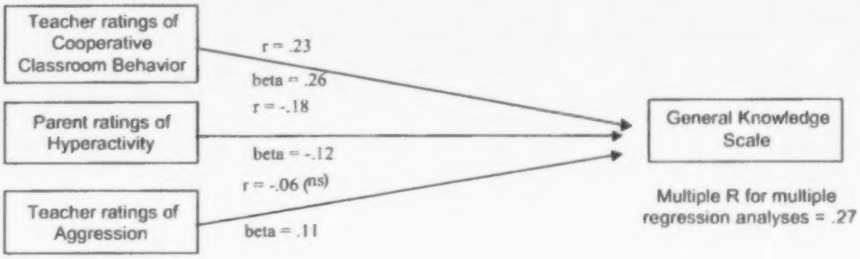
SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 199
Kindergarten Followup, Ns = 741-771 forbivariate correlations, 719 for multiple regression coefficients.

Figure 7.12. Correlation Coefficients Between Parent and Teacher Ratings of Children's Social Competencies at the End of Head Start and General Knowledge Scale Scores at End of Kindergarten



Significant correlations were also found between children's scores on each of the parent and teacher reported behavior ratings at the end of Head Start and General Knowledge scale scores at the end of kindergarten, but these relationships were weaker than those found with Reading scale scores reported above. Significant bivariate correlations with General Knowledge scale scores ranged in absolute value from .23 (for teacher ratings of cooperative classroom behavior) to .09 (for parent reports of aggression), again indicating that the FACES measures have predictive power on outcome criteria at later time points. Significant correlations were in the expected directions. Ratings of problem behaviors were negatively correlated with kindergarten General Knowledge scale scores, suggesting that behaviors that may impede learning are associated with lower skills in the natural sciences and social studies in kindergarten (Figure 7.12).

Figure 7.13. Correlations and Significant Standardized Multiple Regression Coefficients between General Knowledge Scale Scores at End of Kindergarten Year and Parent and Teacher Behavior Ratings at End of Hea Start Year


Teacher ratings of Cooperative Classroom Behavior
Parent ratings of Hyperactivity
Teacher ratings of Aggression
r = .23
beta = .26
r = -.18
beta = -.12
r = -.06 (ns)
beta = .11
General Knowledge Scale
Multiple R for multiple regression analyses = .27

SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 199 Kindergarten Followup. Ns = 742-772 forbivariate correlations, 719 for multiple regression coefficients.

When the behavior ratings were combined in a multiple linear regression model, the model accounted for 7 percent of the variance in General Knowledge scale scores. The strongest predictor in the multiple regression was teacher reported ratings of cooperative classroom behavior (beta = .26), which also showed the strongest bivariate correlation with General Knowledge scale scores, followed by parent reported ratings of hyperactive behavior (beta = -.12). Surprisingly, teacher reported ratings of aggressive behavior had a significant positive regression coefficient (beta = .11). None of the other behavior ratings were significant predictors in the model (Figure 7.13).

### FACES Behavior Ratings Contribute to the Prediction of Reading at the End of Kindergarten

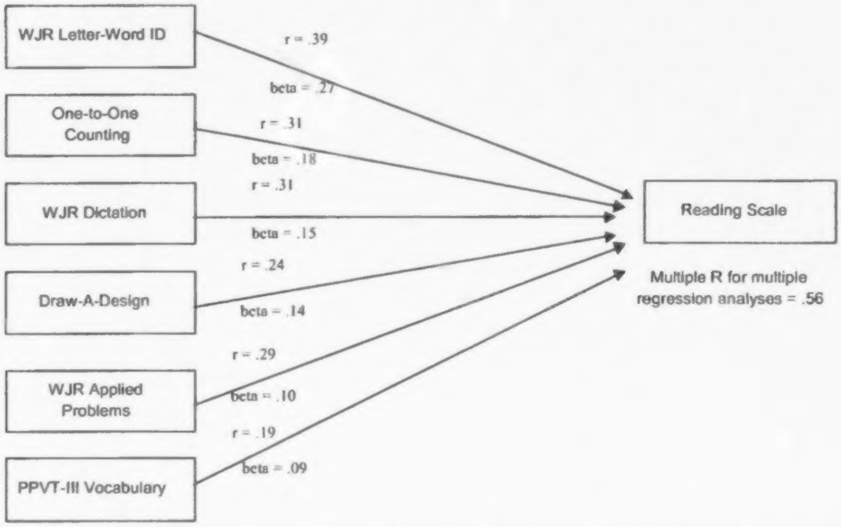
In order to further examine the usefulness of the behavior ratings, the unique contribution of the set of behavior rating measures to the prediction of kindergarten outcomes was assessed. Neither for reading nor for general knowledge did behavior ratings contribute significantly over and above the cognitive assessments ($p > .05$). However, the eight behavior ratings did provide unique contributions at the trend level ($p < .10$) to the prediction of the Reading scale score, over and above the contributions of the cognitive tests, increasing the predictive power of the assessment battery by 1.3 percent. This indicates that the behavior ratings provide some unique contributions to the prediction of kindergarten outcomes.

## D. Head Start Fall to Spring Gain Scores from the FACES Battery Predict Kindergarten Outcomes

As a further assessment of the predictive validity of the cognitive measures from the FACES battery, the predictive validity of their fall to spring gain scores was assessed. To determine the children's gains during their Head Start year, differences between the fall 1997 and spring 1998 mean scores were calculated for each of the measures by subtracting the fall score from the spring score for each child in the study.

The ability of the fall to spring gains scores from the cognitive measures of the FACES battery to predict later school readiness was assessed by two approaches. In the first approach, Reading and General Knowledge scale scores from the end of the kindergarten year were correlated with the gain scores from each of the FACES sub-tests from the end of the Head Start year. In these analyses, partial correlations were calculated, controlling for individual difference in fall 1997 baseline scores.[26] The second approach assessed the ability of the fall to spring gain scores from the FACES instruments to predict Reading and General Knowledge scale scores at the end of the kindergarten year in a multiple linear regression analyses.[27]

Figure 7.14. Correlations and Standardized Multiple Regression Coefficients between Reading Scale Scores at End of Kindergarten Year and FACES Assessment Scale Gain Scores during Head Start Year


WJR Letter-Word ID
r = .39
beta = .27
One-to-One Counting
r = .31
beta = .18
r = .31
WJR Dictation
beta = .15
r = .24
Draw-A-Design
beta = .14
r = .29
WJR Applied Problems
beta = .10
r = .19
beta = .09
PPVT-III Vocabulary
Reading Scale
Multiple R for multiple regression analyses = .56

SOURCE: Head Start Family and Child Experiences Survey (FACES), Fall 1997, Spring 1998 Head Start and Spring 1999 Kindergarten Followup, Ns = 580-697 for bivariate correlations, 441 for multiple regression coefficients.

---

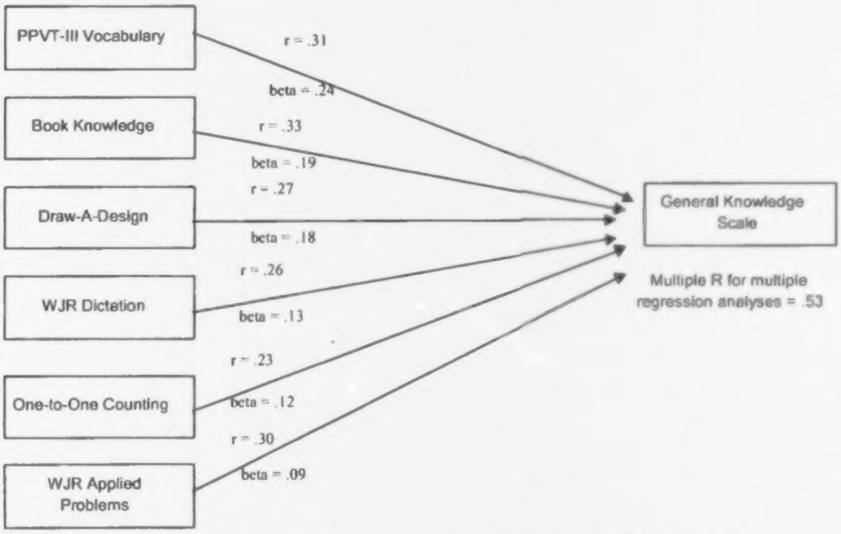
[26] In analyses of gains scores, baseline scores are controlled for, effectively examining the effect of the gain scores, if all students had the same baseline score.

[27] Control of the fall baseline score for each gain score in the multiple linear regression models was accomplished through the use of residual scores. A residual score was created for each gain score with the effects of its respective baseline score partialled out. These residual scores were then entered as independent variables in the multiple linear regression, predicting the kindergarten outcome variables.

Children's fall to spring gain scores on each of the component tasks in the FACES battery correlated significantly with their Reading scale scores at the end of kindergarten. Bivariate correlations with the Reading scale ranged from .39 (for the Woodcock-Johnson-R Letter-Word Identification gain score) to .11 (for the Social Awareness gain score). These significant correlations indicate that the gain scores from the FACES measures are predictive of kindergarten outcomes. When the gain scores were combined in a multiple regression model, the model did quite well at predicting children's early reading skills at the end of kindergarten, accounting for 31 percent of the variance in Reading scale scores. The best predictor of Reading scale scores was the gain score for Woodcock-Johnson-R Letter-Word Identification task (beta = .27), which also showed the highest bivariate correlation with the Reading scale (Figure 7.14).

Similarly, children's fall to spring gain scores on all but one of the component tasks (Color Naming; r = .03) in the FACES battery correlated significantly with their General Knowledge scale scores at the end of kindergarten. Significant bivariate correlations with the General Knowledge scale ranged from .33 (for the Book Knowledge gain score) to .10 (for the Social Awareness gain score). These significant correlations indicate that the gain scores from the FACES measures are predictive of kindergarten outcomes. When the gain scores were combined in a multiple regression model, the model did quite well at predicting children's general knowledge at the end of kindergarten, accounting for 28 percent of the variance in General Knowledge scale scores. The best predictor in the model was the gain score for the PPVT-III (beta = .24), followed by the Book Knowledge gain score (beta = .19), which had the highest bivariate correlation with the General Knowledge scores (Figure 7.15).

Figure 7.15. Correlations and Standardized Multiple Regression Coefficients Between General Knowledge Scale Scores at End of Kindergarten Year and FACES Assessment Scale Gain Scores during Head Start Year



SOURCE: Head Start Family and Child Experiences Survey (FACES), Fall 1997, Spring 1998 Head Start and Spring 1999 Kindergarten Followup, Ns = 580-697 for bivariate correlations, 441 for multiple regression coefficients.

Similar to the analyses with the FACES battery scores at the end of Head Start, the Book Knowledge gain score was not a significant predictor of the Reading scale score in its multiple regression model, though it correlated significantly with the Reading score ($\underline{r} = .26$). Another similarity to the analyses with the scores at the end of Head Start is that the best predictor of the Reading score, namely the Woodcock-Johnson-R Letter-Word Identification gain score, was not a significant predictor of General Knowledge in its multiple regression analysis, despite the fact that its gain score correlated significantly with the General Knowledge task ($\underline{r} = .18$). Conversely, the best predictor of General Knowledge, namely, the PPVT-III, was the weakest significant predictor in the multivariate prediction of children's Reading scores (beta = .09).

The findings of the gain scores analyses support the conclusions of the analyses with the scores at the end of Head Start that the Reading and General Knowledge assessments may be tapping two distinct clusters of skills: "inside-out" skills and "outside-in" skills. The gain scores from the FACES battery again showed its validity by predicting well to children's later learning in both skill domains.

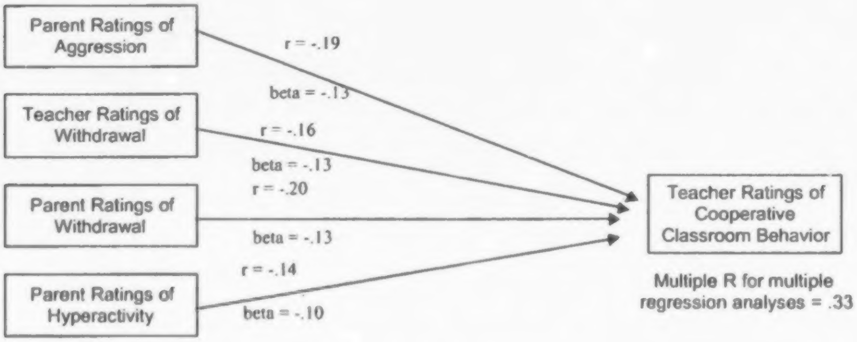
Although both domains were predicted well, the combination of sub-tests that produced the best forecasts differed across the two skill clusters. The gain score from the Letter-Word Identification test was the best predictor of inside-out skills, with gain scores from One-to-One Counting, Dictation, McCarthy Draw-a-Design, Applied Problems, and PPVT-III contributing additional predictive power. The PPVT-III gain score was by far the best predictor of outside-in skills, with gain scores from Book Knowledge, McCarthy Draw-a-Design, Dictation, One-to-One Counting, and Applied Problems showing much smaller but significant regression coefficients as well. It is noteworthy that seven of the eight FACES sub-tests contributed significantly to either the Reading or General Knowledge regression model or both.

### E. Head Start Fall to Spring Gain Scores From the Behavior Ratings Predict Social Competence in Kindergarten

The fall to spring gain scores from the parent and teacher ratings of behavior were moderately correlated with teacher ratings of cooperative classroom behavior at the end of kindergarten, indicating that the FACES measures have predictive power on kindergarten outcomes. Correlations were significant in the expected directions. Increases in problem behaviors as rated by both parents and teachers at the end of Head Start had significantly negative correlations with teacher ratings of cooperative classroom behavior at the end of kindergarten, ranging from -.24 (for the change score for parent reported ratings of total problem behaviors) to -.14 (for the change score for parent reported ratings of hyperactive behavior). The gain score for teacher ratings of cooperative classroom behavior was positively correlated with teacher ratings of cooperative classroom behavior at the end of kindergarten ($\underline{r} = .15$), but the gain score for parent ratings of positive approaches to learning was not. In general, gain score for the teacher ratings showed stronger relationships with the kindergarten outcomes than those for the parent ratings.

When the rating gain scores were combined in a multiple regression, the model accounted for 11 percent of the variance in teacher ratings of cooperative classroom behavior. The best predictors in the multiple regression were the gain scores for teacher reported ratings of withdrawal behavior, and parent reported ratings of aggressive and withdrawal behavior (betas = -.13) which had the stronger bivariate correlations among the set of behavior ratings (Figure 7.16).

Figure 7.16. Correlations and Standardized Multiple Regression Coefficients Between Teacher Ratings of Cooperative Classroom Behavior at End of Kindergarten Year and Parent and Teacher Behavior Ratings Gain Scores during Head Start Year

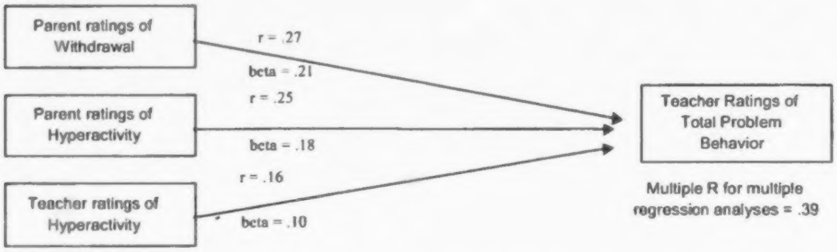

Parent Ratings of Aggression
r = -.19
beta = -.13
Teacher Ratings of Withdrawal
r = -.16
beta = -.13
r = -.20
Parent Ratings of Withdrawal
beta = -.13
r = -.14
Parent Ratings of Hyperactivity
beta = -.10
Teacher Ratings of Cooperative Classroom Behavior
Multiple R for multiple regression analyses = .33

SOURCE: Head Start Family and Child Experiences Survey (FACES), Fall 1997, Spring 1998 Head Start and Spring 1999 KindergartenFollowup, Ns = 474-510 for bivariate correlations, 432 for multiple regression coefficients.

Similarly, the fall to spring gain scores from the parent and teacher ratings of behavior were moderately correlated with teacher ratings of total problem behavior at the end of kindergarten, again indicating that the FACES measures have predictive power on kindergarten outcomes. Correlations were significant in the expected directions. Increases in problem behaviors as rated by both parents and teachers at the end of Head Start had significantly positive correlations with teacher ratings of total problem behavior at the end of kindergarten, ranging from .29 (for the change score for parent reported ratings of total problem behaviors) to .12 (for the change score for teacher reported ratings of aggressive behavior). The gain score for teacher ratings of cooperative classroom behavior was negatively correlated with teacher ratings of cooperative classroom behavior at the end of kindergarten (r = -.13), but the gain score for parent ratings of positive approaches to learning was not significantly related. In general, gain score for the teacher ratings showed stronger relationships with the kindergarten outcomes than those for the parent ratings.

When the rating gain scores were combined in a multiple regression, the model accounted for 15 percent of the variance in teacher ratings of total problem behaviors. The best predictor in the multiple regression were the gain scores for parent reported ratings of withdrawal behavior, (beta = .21) (Figure 7.17).

Figure 7.17. Correlations and Standardized Multiple Regression Coefficients Between Teacher Ratings of Total Problem Behavior at End of Kindergarten Year and Parent and Teacher Behavior Ratings Gain Scores during Head Start Year


Parent ratings of Withdrawal
r = .27
beta = .21
Parent ratings of Hyperactivity
r = .25
beta = .18
Teacher ratings of Hyperactivity
r = .16
beta = .10
Teacher Ratings of Total Problem Behavior
Multiple R for multiple regression analyses = .39

SOURCE: Head Start Family and Child Experiences Survey (FACES), Fall 1997, Spring 1998 Head Start and Spring 1999 Kindergarten Followup , Ns = 474-510 for bivariate correlations, 432 for multiple regression coefficients.

## F. Head Start Fall to Spring Gain Scores From the Behavior Ratings Predict Reading Skills and General Knowledge at the End of Kindergarten

The fall to spring gain scores for parent and teacher ratings of behavior correlated significantly with Reading scale scores at the end of kindergarten, indicating that the FACES measures have predictive power on kindergarten outcomes. Correlations were significant in the expected directions. Increases in problem behaviors were negatively correlated with kindergarten Reading scale scores, adding more evidence that behaviors that may impede learning are associated with lower reading skills in kindergarten. When the behavior ratings gain scores were combined in a multiple regression model, the model accounted for 4 percent of the variance in Reading scale scores. The best predictor in the multiple regression was the gain score for teacher reported ratings of withdrawal behavior (beta = -.12) followed by the gain score for parent reported ratings of withdrawal behavior (beta = -.11).

The correlations of the fall to spring gain scores of parent and teacher ratings of behavior and General Knowledge scale scores at the end of kindergarten were not as strong as those for the Reading scale scores. Four of the eight behavior rating gain scores (parent ratings of aggressive, hyperactive, and withdrawal behavior, and teacher ratings of cooperative classroom behavior) were significantly related to General Knowledge scores, indicating that these FACES measures have some predictive power on kindergarten outcomes. Significant correlations were in the expected directions. Increases in problem behaviors were negatively correlated with kindergarten General Knowledge scale scores, adding more evidence that behaviors that may impede learning are associated with lower skills in natural sciences and social studies in kindergarten. When the behavior ratings were combined in a multiple regression model, the model accounted for 2.5 percent of the variance in General Knowledge scale scores, which is an association at the trend level ($p < .10$).

### Head Start Fall to Spring Gain Scores From the Behavior Ratings Contribute to the Prediction of Reading and General Knowledge at the End of Kindergarten

In order to further examine the usefulness of the behavior ratings, the unique contribution of the set of behavior rating gain scores to the prediction of kindergarten outcomes was assessed. The eight behavior ratings provided unique contributions at the trend level ($p < .10$) to the prediction of the Reading scale score, over and above the contributions of the cognitive tests, increasing the predictive power of the assessment battery by almost 3 percent. Similarly, the eight behavior ratings provided unique contributions at the trend level ($p < .10$) to the prediction of the General Knowledge scale score, over and above the contributions of the cognitive tests, increasing the predictive power of the assessment battery by almost 3 percent. These results indicate that the addition of the behavior rating gain scores provide some unique contributions to the prediction of kindergarten outcomes.

## G. Children's Scores on the FACES Instruments and Behavior Ratings at the End of Head Start Predict Promotion to First Grade

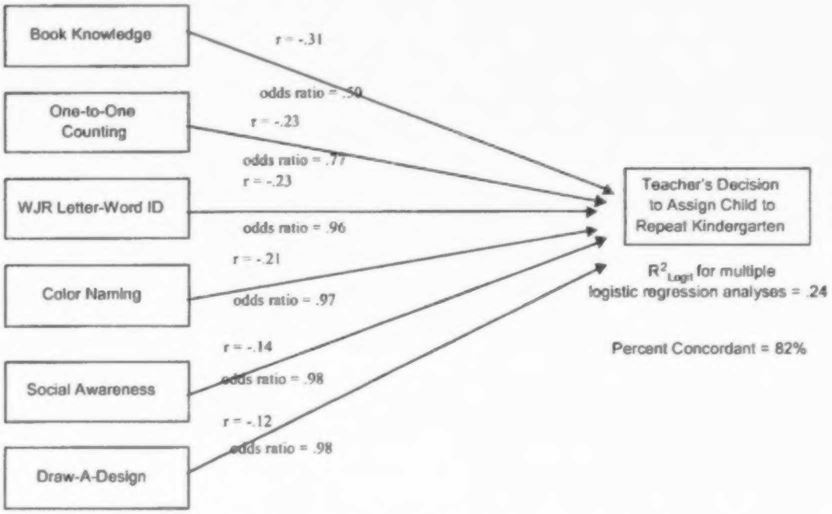
Another measure of the predictive validity of the FACES battery is to examine how well scores on the FACES instruments and behavior ratings at the end of Head Start are related to practical decision-making at the end of kindergarten, namely, the teacher's decision of whether the child gets promoted to first grade or repeats the kindergarten year. In this set of analyses, teachers' decisions at the end of the kindergarten year to have the child repeat another year of kindergarten (versus promote the child to first grade) were first correlated with the scores from each of the FACES instruments from the end of the Head Start year. Then the ability of the FACES scale scores from the end of the Head Start year to predict teachers' decisions to have the child repeat a year of kindergarten was examined in a multiple logistic regression analyses. This approach was then repeated with the parent- and teacher-reported behavior ratings.

### Children's Scores on the FACES Instruments at the End of Head Start Predict Repeating Kindergarten

Children's scores on each of the component tasks in the FACES battery at the end of Head Start correlated significantly with teacher decisions to have the child repeat another year of kindergarten. Bivariate correlations were in the expected directions, ranging from -.31 (for Book Knowledge) to -.12 (for Draw-A-Design), indicating that lower subtest scores were associated with repeating kindergarten. These significant correlations indicate that the FACES measures have predictive power on outcome criteria at later time points.

When the subtest scores were combined in a multiple logistic regression model, the model did quite well at predicting whether children repeated kindergarten, accounting for 24 percent of the variance in the prediction of repeating kindergarten. Information of scores from the FACES instruments led to an 82 percent accuracy rate in predicting or not a child was assigned by her teacher to repeat kindergarten. The strongest predictor of whether or not children were assigned by their teachers to repeat kindergarten was the Book Knowledge task, in which for every unit increase in Book Knowledge scores, children were 50 percent less likely to repeat kindergarten (Figure 7.18).

Figure 7.18. Correlations and Odds-Ratio Estimates Between Teachers' Decisions at End of Kindergarten Year to Assign Child to Repeat Kindergarten and FACES Assessment Scale Scores at end of Head Start Year


Book Knowledge
r = -.31
odds ratio = .50
One-to-One Counting
r = -.23
odds ratio = .77
WJR Letter-Word ID
r = -.23
odds ratio = .96
Color Naming
r = -.21
odds ratio = .97
Social Awareness
r = -.14
odds ratio = .98
Draw-A-Design
r = -.12
odds ratio = .98
Teacher's Decision to Assign Child to Repeat Kindergarten
$R^2_{Logit}$ for multiple logistic regression analyses = .24
Percent Concordant = 82%

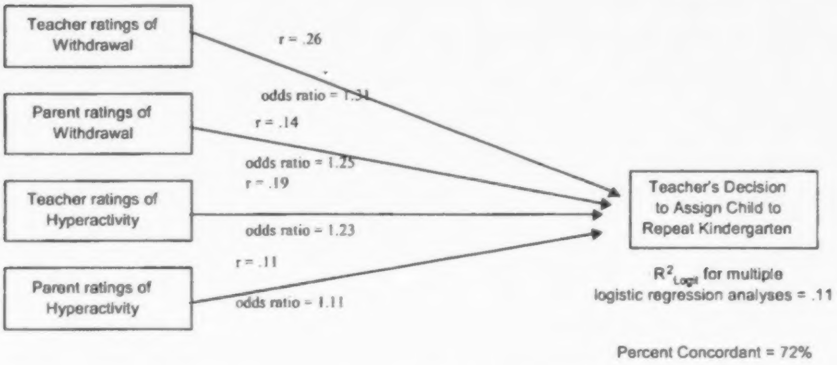
SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 1999 Kindergarten Followup, Ns = 551-583 for bivariate correlations, 474 for multiple logistic regression odds ratio estimates.

## Children's Behavior Ratings at the End of Head Start Predict Repeating Kindergarten

Parent and teacher behavior ratings at the end of Head Start correlated significantly with teacher decisions to have the child repeat another year of kindergarten. Correlations were in the expected directions. Problematic behaviors were positively correlated with repeating kindergarten, ranging from .26 (for teacher ratings of withdrawal behavior) to -.09 (for teacher ratings of aggressive behavior). Teacher ratings of cooperative classroom behavior at the end of Head Start was negatively correlated with teacher decisions to repeat kindergarten (r = -.14); parent ratings of positive approaches to learning, however, was not significantly related. The significant correlations indicate that the FACES behavior ratings also have predictive power on outcome criteria at later time points.

When the parent and teacher ratings were combined in a multiple logistic regression model, the model did quite well at predicting whether children repeated kindergarten, accounting for 11 percent of the variance in the prediction of repeating kindergarten. Information from the FACES behavior ratings led to a 72 percent accuracy rate in predicting or not a child was assigned by her teacher to repeat kindergarten. The strongest predictor of whether or not children were assigned by their teachers to repeat kindergarten was teacher ratings of withdrawal behavior, in which for every unit increase in these teacher ratings, children were 31 percent more likely to repeat kindergarten (Figure 7.19).

Figure 7.19. Correlations and Odds-Ratio Estimates Between Teachers' Decisions at End of Kindergarten Year to Assign Child to Repeat Kindergarten and Parent and Teacher Behavior Ratings at end of Head Start Year


Teacher ratings of Withdrawal
r = .26
odds ratio = 1.31
Parent ratings of Withdrawal
r = .14
odds ratio = 1.25
Teacher ratings of Hyperactivity
r = .19
odds ratio = 1.23
Parent ratings of Hyperactivity
r = .11
odds ratio = 1.11
Teacher's Decision to Assign Child to Repeat Kindergarten
R²Logit for multiple logistic regression analyses = .11
Percent Concordant = 72%

SOURCE: Head Start Family and Child Experiences Survey (FACES), Spring 1998 Head Start and Spring 1999 Kindergarten Followup, Ns = 551-583 for bivariate correlations, 474 for multiple logistic regression odds ratio estimates.

## The Combination of Children's Scores on the FACES Instruments and Behavior Ratings at the End of Head Start Predict Repeating Kindergarten

When the parent and teacher ratings were combined with the subtest scores from the FACES instruments in a multiple logistic regression model, the model did quite well at predicting whether children repeated kindergarten, accounting for 30 percent of the variance in the prediction of repeating kindergarten. This is an increase of 6 percent of the variance that was explained by the subtest scores alone. Information from the combination of the FACES behavior ratings and the subtest scores led to an 83 percent accuracy rate in predicting whether or not a child was assigned by her teacher to repeat kindergarten. This suggests that the additional information provided by the behavior ratings adds to the predictive validity of the FACES instruments in predicting kindergarten repetition.

## CONCLUSIONS

The FACES battery has strong predictive validity with outcomes at the end of kindergarten. As an indicator of preliteracy skills, the cognitive measures show strong associations with reading ability at the end of the kindergarten year. As an indicator of school adjustment and social competence, the behavior ratings demonstrate ability to predict kindergarten behaviors that promote learning. These analyses show that:

- The instruments used in FACES predict later behavior and performance in Kindergarten.
- The instruments used in FACES also predict the later practical decision of whether a child gets promoted to first grade.
- The instruments used in FACES tap different types of abilities ("inside-out" versus "outside-in") that are important for children's future reading proficiency and academic achievement.
- The multi-measure and multi-method approach to the measurement of children's abilities provides a variety of information sources that significantly contribute to the prediction of kindergarten outcomes.

Children who had higher scores at the end of the Head Start year, and who made greater gains during the year on the Letter-Word Identification test, Applied Problems, Dictation, One-to-One Counting, and McCarthy Draw-a-Design tasks, tended to have greater early reading skills at the end of kindergarten. Children's improved scores on the PPVT-III, Applied Problems, Book Knowledge, Dictation, and the McCarthy Draw-a-Design tasks at the end of Head Start were associated with greater General Knowledge scores at the end of kindergarten. These results suggest that efforts in improving children's performance and behavior in preschool can result in greater school readiness and school adjustment when these children are preparing to enter first grade.

In the assessment of children's social competencies, the use of parent and teacher ratings provides data on children's coping skills in different situations and provides a more comprehensive picture of their behavior. Equally important, both parent and teacher ratings significantly contribute to the prediction of social skills at the end of kindergarten. The parent and teacher ratings also significantly predict reading skills and general knowledge at the end of kindergarten. Ratings of problem behaviors were negatively correlated with kindergarten Reading and General Knowledge scale scores, suggesting that behaviors that may impede learning are associated with lower reading skills in kindergarten. High ratings of behaviors that enhance learning, positive approaches to learning and cooperative classroom behavior, were positively correlated with kindergarten outcomes. These analyses suggest that curricula that strengthen children's social skills will also have beneficial effects on their later school readiness.

The multi-measure and multi-method approach to the measurement of children's development and school readiness provides a comprehensive assessment of children's abilities. The addition of the criterion-referenced measures to the norm-referenced measures improves the assessment battery in many ways. First, they are short tasks that cover more specific topic areas that are typically taught in preschool curricula and are also fun for the children to do in the assessment. And second, they significantly (although moderately) increase the battery's ability to predict kindergarten outcomes, improving its predictive validity. The addition of the parent and teacher behavior ratings adds another source of information that predict kindergarten outcomes, namely an assessment of behaviors that can either foster or impede learning. They also provide some unique contributions to the battery's ability to predict kindergarten outcomes, particularly in the practical decision of whether a child is promoted to first grade or repeats kindergarten.

## REFERENCE

Whitehurst, G. J. & Lonigan, C. J. (1998) Child development and emergent literacy. *Child Development, 69*, 848-872.

## References

Abbott-Shim, M. & Sibley, A. (1998). *Assessment Profile for Early Childhood Programs: Research version.* Atlanta, GA: Quality Assist.

Abbott-Shim, M, Lambert, R, & McCarty, F. (2000). Structural model of Head Start classroom quality. *Early Childhood Research Quarterly*, 15, 1, 115-134.

Administration on Children, Youth and Families. (2002). *A descriptive study of Head Start families: FACES technical report I.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth and Families. (2001). *Head Start FACES: Longitudinal findings on program performance. Third progress report.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth and Families. (1998). *Head Start Program Performance Measures: Second progress report.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth and Families. (1997). *First progress report on the Head Start Program Performance Measures.* Washington, DC: U.S. Department of Health and Human Services.

Administration on Children, Youth and Families. (1995). *Charting our progress: Development of the Head Start Program Performance Measures.* Washington, DC: U.S. Department of Health and Human Services.

Annie E. Casey Foundation (1999). *1999 Kids count data book: State profiles of child well-being.* Baltimore, MD: Annie E. Casey Foundation.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-552.

Barnett, W.S. (1998). Long-term effects on cognitive development and school success. Pp. 11-44 in *Early Care and Education for Children in Poverty*, edited by W.S. Barnett & S.S. Boocock. Albany, NY: State University of New York Press.

Blau, D. (1997). The production of quality in child care centers. *Journal of Human Resources*, 32, 354-387.

Bredekamp, S. (1986). The reliability and validity of the early childhood classroom observation scale for accrediting early childhood programs. *Early Childhood Research Quarterly*, 1, 103-118.

Bryant, D. M., Burchinal, M., Lau, L. B., and Sparling, J.J. (1994). Family and classroom correlates of Head Start children's developmental outcomes. *Early Childhood Research Quarterly, 9*, 289-309.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Burts, D.C., C.H. Hart, R. Charlesworth, & Kirk, L. (1990). A comparison of frequencies of stress behaviors observed in kindergarten children in classrooms with developmentally appropriate versus developmentally inappropriate practices. *Early Childhood Research Quarterly*, 5, 407-423.

Cost, Quality & Child Outcomes Study Team. (1995). *Cost, quality and child outcomes in child care centers, public report, second edition.* Denver: Economics Department, University of Colorado at Denver.

Dunn, L.M., Padilla, E.R., Lugo, D.E., & Dunn, L.M. (1986). *Test de Vocabulario en Imagenes Peabody.* Circle Pines, MN: American Guidance Service.

Feldhaus, K.M., Koziol-McLain, J., & Amsbury, H.L. (1997). Accuracy of 3 brief screening questions for detecting partner violence in the emergency department. *Journal of the American Medical Association,* 227(17), 1357.

Fishel, J.E., Storch, S.A., Spira, E.G., & Stolz, B.M. (2003). Enhancing emergent literacy skills in Head Start: First year curriculum evaluation results. Presented at Biennial Meeting of the Society for Research in Child Development, Tampa, FL.

Goal One Technical Planning Group. (1991). The Goal One Technical Planning Subgroup report on school readiness. In National Education Goals Panel (Ed.), *Potential strategies for long-term indicator development: Reports of the technical planning subgroups* (Report No. 91-0, pp. 1-18). Washington, DC: National Education Goals Panel.

Goal One Technical Planning Group. (1993). *Reconsidering children's early development and learning: Toward shared beliefs and vocabulary. Draft report to the National Education Goals Panel.* Washington, DC: National Education Goals Panel.

Hall, L., Williams, C., & Greenberg, R. (1985). Supports, stressors, and depressive symptoms in low-income mothers of young children. *American Journal of Public Health, 75*, 518-522.

Haskins, R. (1989). Beyond metaphor: The efficacy of early childhood education. *American Psychologist,* 44(2), 274-282.

Hayes, C., Palmer, J., & Zaslow, M. (1990). *Who cares for America's children? Child care policy for the 1990s.* Report of the Panel on Child Care Policy, Committee on Child Development Research and Public Policy, National Research Council. Washington, DC: National Academy Press.

Howes, C., Phillips, D., & Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based child care. *Child Development,* 63, 449-460.

Kontos, S., and Fiene, R. (1987). Child care quality, compliance with regulations, and children's development: The Pennyslvania Study. In D. Phillips (ed.), *Quality in child care: What does the research tell us?* (pp. 57-79). Washington, DC: National Association for the Education of Young Children.

Lange, S.M. (2000). The challenges confronting children of incarcerated parents. *Journal of Family Psychotherapy, 11(4)*, 61-68.

Layzer, J., Goodson., B., & Moss, M. (1993). *Final report volume 1: Life in preschool.* Cambridge, MA: Abt Associates, Inc.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126(2)*, 309-337.

Liaw, F., & Brooks-Gunn, J. (1994). Cumulative familial risks and low birth weight children's cognitive and behavioral development. *Journal of Clinical Child Psychology, 23*, 360-372.

McKey, R.H., Condelli, L., Ganson, H., Barrett, B.J., McConkey, C., & Plantz, M.C. (1985). *The impact of Head Start on children, families, and communities.* (DHHS Publication No. OHDS 85-31193). Washington, DC: U.S. Government Printing Office.

NICHD Early Child Care Research Network (ECCRN). (2000). The relation of child care to cognitive and language development. *Child Development, 71*, 823-839.

Peisner-Feinberg, E. S. and Burchinal, M. R. (1997). Relations between preschool children's child-care experiences and concurrent development: The Cost, Quality and Outcomes Study. *Merrill-Palmer Quarterly.* 43, 3. 451-477.

Peisner-Feinberg, E.S., Burchinal, M.R., Clifford, R.M., Culkin, M.L, Howes, C., Kagan, S.L., and Yazejian, N. (2001). The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Developmental Trajectories through Second Grade. *Child Development, 72(5)*, 1534-1553.

Phillips, D. & Howes, C. (1987). *Indicators of quality in child care: Review of the research.* In D. A. Phillips (Ed.), Quality in child care: What does research tell us? (pp. 1-19). Research Monograph of the National Association for the Education Young Children, Vol. 1. Washington, DC: NAEYC.

Phillips, D., McCartney, K., and Scarr, S. (1987). Child care quality and children's social development. *Developmental Psychology, 23*, 537-543.

Phillips, D., Mekos, D., Scarr, S., McCartney, K. & Abbott-Shim, M. (2000). Within and beyond the classroom door: Assessing quality in child care centers. *Early Childhood Research Quarterly.* 15, 4. 475-496.

Phillips, M., Brooks-Gunn, J., Duncan, G.J., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the black-white test score gap. Pp. 103-145 in *The Black-White Test Score Gap*, edited by C. Jencks & M. Phillips, Washington DC: Brookings Institution Press.

Pianta, R.C., & McCoy, S.J. (1997). The first day of school: The predictive validity of early school screening. *Journal of Applied Developmental Psychology, 18*, 1-22.

Radloff, L.S. (1977). The CES-D scale: A self report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.

Resnick, G., McKey, R.H., & Klayman, D. (2001). *The evaluation of early childhood education programming in the 30 Abbott School Districts: First-year report on program implementation and descriptions of children and families*. (Contract No. A43078). Trenton, NJ: New Jersey Department of Human Services and New Jersey Department of Education.

Rosenthal, R. & Rosnow, R.L. (1984). *Essentials of behavioral analysis: Methods and data analysis.* New York: McGraw-Hill.

Ruopp, R., Travers, J., Glantz, F., & Coelen, C. (1979). *Children at the center: Final results of the National Day Care Study.* Boston: Abt Associates.

Scarr, S., & Eisenberg, M. (1993). Child care research: Issues, perspectives and results. *Annual Review of Psychology,* 44, 613-644.

Scarr, S., M. Eisenberg, & Deater-Deckard, K. (1992). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, 9, 131-151.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24(4), 323-355.

Smith, T., Kleiner, A., Parsad, B., & Farris, E. (2002). *Prekindergarten in U.S. public schools: 2000–2001* (NCES 2003-019). U.S. Department of Education, National Center for Education Statistics: Washington, DC.

Whitebook, M., Howes, C., & Phillips, D. (1989). *Who Cares? Child care teachers and the quality of care in America* (Final report of the National Child Care Staffing Study). Oakland, CA: Child Care Employee Project.

Whitehurst, G.J., & Lonigan, C.J. (1998). Child development and emergent literacy. *Child Development,* 69, 848-872.

Zill, N., & West, J. (2001). *Entering kindergarten: A portrait of American children when they begin school.* Washington, DC: National Center for Education Statistics, U.S. Department of Education.

# Head Start FACES 2000:
# A Whole-Child Perspective on Program Performance

## Appendix

## Methodology

## I. Sample Design Overview

The third cohort of Head Start children for FACES was selected as a two-stage sample.[28] The first stage sampling units were Head Start programs; the second stage units were classes within sampled programs. In each sampled classroom, all eligible children in their first year of Head Start were taken into the sample.

### A. Programs

The sampling frame of eligible Head Start programs was constructed from the 1998-1999 Program Information Report (PIR). Migrant and Seasonal Head Start programs, American Indian/Alaska Native Head Start programs, Early Head Start programs, programs in the territories, and programs that do not serve children directly were excluded, resulting in a frame of 1,675 programs. The programs were stratified by Census region (NE, NC, S, W), percent minority (above/below 50%), and metro or urban/rural status (MSA/non-MSA). These are the same stratification variables used in sampling programs for the first (Spring 1997) and second (Fall 1997 – Spring 2000) FACES cohorts.

A sample of 45 programs was selected in spring 2000. The sample size in each stratum was proportional to the stratum first year Head Start enrollment. The programs were selected with probability proportional to the program's first year enrollment using systematic sampling. The first year enrollment was calculated from the PIR by subtracting the reported second and third year enrollment from the total enrollment. A Keyfitz procedure was used to minimize the overlap with the 40 programs sampled previously by Abt Associates for the first and second FACES cohorts. As a result, there was no overlap with the previous program sample. Of the 45 programs selected for the third cohort, two were later discovered to be ineligible because they had been defunded.

---

[28] The first cohort consisted of children sampled for the Spring, 1997 Field Test. The second cohort consisted of an additional sample of children selected for the FACES study that commenced in Fall, 1997. The combination of these two cohorts comprised the total sample for the FACES 1997 study. The third cohort consists of an entirely new sample of children selected for the current FACES study that commenced in Fall, 2000.

## B. Classrooms

In the 43 remaining programs, lists of the anticipated classes for fall 2000 were obtained in late summer 2000. The programs also provided the expected number of first year Head Start children in each class. These lists formed the basis for the classroom sampling frame, after excluding classes with no first year children. Classes with fewer than five first year children expected were combined with another class in the same center to form a "class group." The class groups were treated as a single unit for sampling purposes and sample size calculations. The total target sample size of first year children was 2825, or 66 per program. In general, the desired sample size of classes in each program was determined as 66 / (average class size for the program), where the average class size was in terms of number of first year children. The actual initial sample size was increased by 2 classes to allow for a reserve sample in each program. In programs where the total first year enrollment as obtained from the class rosters was more than twice the measure of size used to sample the program, the initial class sample size was increased to prevent large variation in class weights. In small programs where the initial sample size exceeded the number of classes available, all classes were taken with certainty.

Classes were sorted by center within program and were sampled with equal probabilities. A subsample of the initial sample was selected with equal probabilities to obtain a main sample of the desired sample size and a reserve sample of two classes in each program. A total of 367 classes were selected: 279 classes for the main sample and 88 for the reserve sample. The number of main sample classes in each program varied from 3 to 15, with an average of 6. (In terms of collapsed classrooms, a total of 252 classroom groups were sampled for the main sample and 82 for the reserve sample, for an average of 6 per program and a range of 3 to 10.)

In Fall 2000 the eligibility status of the main sample classes was determined. One or two reserve classes were added in some programs to prevent a shortfall in the target number of first year children for the study. The final sample for weighting purposes included all sampled classes where an attempt was made to collect data from the classroom, including those discovered to be ineligible. The rationale for this is because ineligible classes in the sample represent ineligible classes on the Head Start program frame. A total of 307 main and reserve classes were in sample in Fall 2000. Twenty of the 307 classes were discovered to be ineligible when the program was contacted by field staff in Fall 2000 because they no longer existed, they didn't receive Head Start funding, or they had no first year Head Start children. In 286 of the remaining 287 eligible classes (one teacher refused to allow the children in her class to be sampled), all first year children were taken into the sample.

## II. Response Rates

### Fall 2000

- 2,508 child assessments were completed out of 2,790 for a completion rate of 90 percent.
- 2,488 parent interviews were completed out of 2,790 families selected for the sample (89 percent).
- Teacher report forms were obtained on 2,532 of the sample children (91 percent).
- Assessment, parent, and teacher data were obtained on 2,396 of the 2,790 sample children (86 percent).
- A total of 278 classrooms were observed out of 286 in the sample for a completion rate of 97 percent.

### Spring 2001

- 2,232 child assessments were completed out of 2,288, representing 98 percent of the children who remained in the program, and 80 percent of the original sample (2,790).
- 2,166 parent interviews were completed out of 2,288, representing 95 percent of the children who remained in the program, and 78 percent of the original sample.
- Teacher report forms were obtained on 2,236 of the sample children, representing 98 percent of the children who remained in the program and 80 percent of the original sample.
- Assessment, parent, and teacher data were obtained on 2,115 of the 2,288 sample children who remained in the program (92 percent).
- A total of 275 classrooms were observed out of 284 in the sample for a completion rate of 97 percent.

### Spring 2002 (Kindergartners Only)

- 831 child assessments were completed out of 979, representing 85 percent of the children who were in Kindergarten in Spring, 2002.
- 901 parent interviews were completed out of 979, representing 92 percent of the children who were in Kindergarten in Spring, 2002.
- Teacher report forms were obtained on 681 of the children, representing 70 percent of the children who were in Kindergarten in Spring, 2002.
- Assessment, parent, and teacher data were obtained on 624 of the 979 children who were in Kindergarten in Spring, 2002 (64%).

## III. Program Weights

The program weight was calculated as the inverse of the program's probability of selection. As mentioned earlier, a Keyfitz procedure was used to minimize the overlap with the cohort 1 and 2 program sample drawn earlier by Abt Associates. This procedure involved calculating conditional probabilities of selection which are based on whether the program was sampled previously or not, and whether it's probability of selection increased compared with the previous sample. Prior to sampling for cohort 3, the unconditional probability of selection for each program on the cohort 3 frame was calculated as

$$\mathrm{NEWPSEL}_i = \frac{\mathrm{FIRSTYR}_i}{(\sum_{i=1}^{Nh} \mathrm{FIRSTYR}_i)/\mathrm{NEWSMPSZ}_h}$$

where $N_h$ is the number of programs on the frame in stratum h, $\mathrm{NEWSMPSZ}_h$ is the sample size for stratum h for the cohort 3 design, and $\mathrm{FIRSTYR}_i$ is the first year enrollment for program i from the PIR. The probability of selection for each program under the Abt sample design for Cohorts 1 and 2 was also calculated as

$$\mathrm{ORIGPSEL}_i = \frac{\mathrm{ENRTOT}_i}{(\sum_{i=1}^{Nh} \mathrm{ENRTOT}_i)/\mathrm{SAMPSIZE}_h}$$

where $N_h$ was the number of programs on the Abt frame in stratum h, $\mathrm{SAMPSIZE}_h$ was the sample size for stratum h under the Abt design, and $\mathrm{ENRTOT}_i$ was the total enrollment for the i-th program from an earlier PIR.

The conditional probability of selection was calculated for each program on the cohort 3 frame according to the Keyfitz procedure as:

Case 1: NEWPSEL $\geq$ 1 – ORIGPSEL

$\mathrm{CONDPROB} = \frac{\mathrm{NEWPSEL} - (1 - \mathrm{ORIGPSEL})}{\mathrm{ORIGPSEL}}$ if the program was sampled for cohorts 1 and 2,

= 1 if the program was not sampled for cohorts 1 and 2.

Case 2: NEWPSEL < 1 – ORIGPSEL

$$\text{CONDPROB} \begin{cases} = 0 & \text{if the program was sampled for cohorts 1 and 2,} \\ = \dfrac{\text{NEWPSEL}}{1-\text{ORIGPSEL}} & \text{if the program was not sampled for cohorts 1 and 2.} \end{cases}$$

These conditional probabilities of selection were the measures of size used to select the cohort 3 program sample. It can be shown that the Keyfitz procedure preserves the unconditional cohort 3 program probabilities of selection, while at the same time minimizing the overlap. Thus the cohort 3 program weight is the inverse of NEWPSEL, the unconditional probability of selection under the cohort 3 design.

All 43 eligible programs cooperated with the study, so that nonresponse adjustments at the program level were unnecessary.

For each program, a set of 43 jackknife replicate weights was created for calculating standard errors. The replicate weights were created using a standard stratified jackknife procedure. One program at a time was dropped (i.e. given a zero replicate weight) and the weights of the remaining programs in the same stratum were adjusted by a factor of nh/(nh-1), where nh is the number of sampled programs in stratum h. The program weights in the other strata were left unchanged. By repeating this 43 times, 43 replicate weights were obtained for each program. For estimates involving child or classroom data from all 43 programs, the degrees of freedom for the variance of the estimate is #PSUs - #varstrat = 43 – 12 = 31. (One of the 13 original sampling strata was collapsed with an adjacent stratum for variance estimation purposes because it contained only one eligible sampled program.)

### A. Classroom Weighting

Two sets of class weights were produced for classroom level estimation: one set for Fall 2000 cross-sectional estimates and a second set for Fall 2000 – Spring 2001 longitudinal classroom analysis. Class base weights were first created that reflected the overall probability of selection for the class, including the program probability of selection. These base weights were adjusted for classroom level nonresponse, using the following criteria for a complete classroom:

Fall 2000 cross-sectional estimates: the classroom must have complete Fall 2000 observation data. Classroom observation data includes counts of children and adults, Assessment Profile (Scheduling, Learning Environment, and Individualizing), ECERS-R, Arnett Caregiver Interaction Scale, Teacher-Directed Activities Checklist and Wrap-Up measures.

Fall 2000-Spring 2001 longitudinal analysis: the classroom must have complete observation data for either Fall 2000 or Spring 2001 and child assessment data for both Fall 2000 and Spring 2001.

### A1. Class Base Weights

A class base weight was created for each of the 367 initially sampled classes in Fall 2000. Fifty-four reserve classes that were never used were given base weights of zero. Six main sample classes were sampled out on an ad hoc basis by field staff to reduce burden and to have independence between classes. They were assigned base weights of zero, since they were not part of the final sample. In this situation, a teacher had both a morning and afternoon class in the sample. One class out of the morning/afternoon pair was subsampled.

The remaining 307 classes considered to constitute the sample were each assigned a class base weight equal to the inverse of their overall probability of selection. The overall probability of selection is the product of the program probability of selection and the probability of selecting the class within the program. The inverse of the overall probability of selection can also be written as the product of the program weight and the within-program class weight:

> Class Base Weight = Program Weight * (Total # Classes in Program / # sampled classes fielded)

Collapsed classrooms were counted as one classroom in the base weight calculations, since they were treated as a single unit in sampling. The ad hoc subsampling was reflected by multiplying the base weight of the retained class in the am/pm pair by a factor of 2 and the dropped class by zero. One class that had merged with another was given a zero base weight, and the newly merged class had its base weight multiplied by a factor of .5 to reflect its increased probability of selection.

Forty-three jackknife class replicate base weights were created from the program replicate weights:

> Class Replicate Base Weight j = Program Replicate Weight j * (Total #Classes in Program / #sampled classes fielded); j = 1, 2, …43.

## A2. Cross-sectional Fall 2000 Class Weights

Of the 307 sampled classes that were fielded in Fall 2000, 279 were eligible and had complete classroom data, 8 were eligible but didn't complete data collection, and 20 were discovered to be ineligible. A class nonresponse adjustment factor was applied to the class base weights of the 279. The nonresponse adjustment factor was computed separately by program. Both the 8 incomplete and 20 ineligible classes were given a zero final class weight. The classroom replicate base weights were also adjusted for nonresponse by program, so that the sampling variability in the nonresponse adjustments will be reflected in the standard error estimates.

The sum of the nonresponse-adjusted Fall 2000 classroom weights is 34,638. The unweighted and weighted completion rates are both 97%, excluding ineligibles from both numerator and denominator. The unweighted and weighted eligibility rates are both 94%. The class base weight was used in calculating the weighted rates.

### A3. Longitudinal Fall 2000 – Spring 2001 Class Weights

Of the 286 eligible classes in Fall 2000, 280 completed data collection in Spring 2001. Note that the 279 Fall 2000 classroom completes are not a subset of the 280 Spring 2001 completes. Five classes that completed Fall 2000 data collection did not complete the Spring 2001, and six classes that completed Spring 2001 data collection did not complete the Fall 2000. There were 79 new classes added in Spring 2001 because children who switched classes after the Fall 2000 data collection were followed to the new class. However, no classroom observations were done at these new classes, so they were not considered to be part of the classroom sample and were assigned a zero base weight.

A class nonresponse adjustment factor was applied to the class base weights of the 280 eligible completes. The nonresponse adjustment factor was computed separately by program. The classroom replicate base weights were also adjusted for nonresponse by program, so that the sampling variability in the nonresponse adjustments will be reflected in the standard error estimates. The incomplete and ineligible classes, along with the 79 new classes, were given a zero final class weight.

The sum of the nonresponse-adjusted Fall 2000-Spring 2001 classroom weights is 34,768. The unweighted and weighted completion rates are both 98%, excluding ineligibles from both numerator and denominator. Both unweighted and weighted eligibility rates are 94%. The class base weight was used in calculating the weighted rates.

### B. Child Weights

Two sets of child weights were produced: a cross-sectional set for Fall 2000 estimates, and a Fall 2000 – Spring 2001 set for longitudinal analyses. Child base weights were first created that reflected the overall probability of selection for the child, including the program and classroom stages of sampling. These base weights were adjusted for child nonresponse, using the following criteria for a complete child case:

Fall 2000 cross-sectional analysis: a child is considered a complete case if the child has a parent interview from either Fall 2000 or Spring 2001, and a Fall 2000 child assessment or teacher rating.

Fall 2000-Spring 2001 longitudinal analysis: a child is considered a complete case if the child has either a Fall 2000 or Spring 2001 parent interview, and one of the following data pairs: a child assessment for both Fall 2000 and Spring 2001, or a teacher rating for both Fall 2000 and Spring 2001.

### B1. Child Base Weights

In 286 eligible Fall 2000 classes, all eligible children in their first year of Head Start were taken into the sample with certainty. A base weight was created for each child as the product of their program weight and nonresponse-adjusted classroom weight. Note that these nonresponse adjusted class weights are not the same as those described earlier, which were designed for use in classroom level analyses. The creation of special classroom weights for the child weights was necessary because there were eligible classrooms that did not have complete classroom observations, but did allow their children to be sampled, and vice versa. To create this special classroom weight, the classroom base weight was adjusted for classes which had eligible children but where "sampling" of children did not take place. This nonresponse-adjusted classroom weight was then used in calculating the child base weight. Since there was no subsampling of children within classrooms, the within-classroom child weight is equal to one and the overall child weight can be written as:

Child Base Weight = Program Weight * Nonresponse-adjusted Classroom Weight.

A set of 43 jackknife (JKn) replicate base weights was also created for each child using the program replicate weights and the special full-sample nonresponse-adjusted classroom weight:

Child Replicate Base Weight j = Program Replicate Weight j * Nonresponse-adjusted Classroom Weight; j = 1, 2, ...43.

### B2. Child Fall 2000 Cross-Sectional Weights

Of the 3,100 children in the Fall 2000 sample, 2,535 were considered completes for the Fall 2000 data collection, 251 were eligible but incomplete (30 of these had assessments but no parent interview), and 314 were ineligible. Children could be ineligible if they came from classrooms that were ineligible, or they were discovered to be in their second year of Head Start, or were otherwise ineligible when Fall 2000 data collection began.

The child base weights of the eligible, complete children in each classroom were adjusted for nonresponse separately by classroom. The ineligible and incomplete children were given a zero final child weight and were dropped from the sample for the Spring 2001 data collection. The replicate child base weights were also adjusted for nonresponse by classroom, so that the sampling variability in the nonresponse adjustments will be reflected in the standard error estimates.

The sum of the nonresponse-adjusted Fall 2000 child weights is 337,247. The unweighted and weighted completion rates are both 91%, excluding ineligibles from both the numerator and denominator. The unweighted and weighted eligibility rates are 90% and 91%, respectively. The child base weight was used in calculating the weighted rates.

### B3. Child Fall 2000-Spring 2001 Longitudinal Weights

In Spring 2001 the eligible first year children were again given assessments, a teacher rating, and an attempt was made to interview the child's parent(s). Of the 2,535 eligible children who had completed Fall 2000 data collection, 2,359 were eligible, complete cases for the Fall 2000 – Spring 2001 data collection; 171 were eligible, incompletes; and five became ineligible because they moved out of the area.

Children who had switched to new classes in the Spring 2001 were followed up, but classroom observations were not done at the new classes. There were 91 children from the Fall 2000 sample who were followed to 79 new classrooms in Spring 2001. In calculating their base weights, these children were given the classroom probability of selection associated with the classroom from which they were originally sampled in Fall 2000.

The child base weights of the eligible, complete children in each classroom were adjusted for nonresponse separately by classroom. The ineligible and incomplete children were given a zero final child weight. The replicate child base weights were also adjusted for nonresponse by classroom, so that the sampling variability in the nonresponse adjustments will be reflected in the standard error estimates.

The sum of the nonresponse-adjusted Fall 2000-Spring 2001 child weights is 338,047. The unweighted and weighted conditional Spring 2001 completion rates are both 93%. The conditional rate is the percent of Fall 2000 eligible completes who also completed the Spring 2001 data collection. The overall (unconditional) completion rate is the product of the completion rates for the Fall 2000 and Spring 2001 data collections: 91% * 93% = 85%. This rate is the percent of eligible, sampled children in Fall 2000 who completed the Spring 2001 data collection.

## IV. Data Collection Instruments

### A. Direct Child Assessment

### A1. Peabody Picture Vocabulary Test – Third Edition - Revised

The Peabody Picture Vocabulary Test (PPVT-III) (Dunn & Dunn, 1997) is designed to assess children's knowledge of the meaning of words by asking them to say or indicate by pointing which of four pictures best shows the meaning of a word that is said aloud by the assessor. A series of words is presented, ranging from easy to difficult for children of a given age, each accompanied by a picture plate consisting of four line drawings. The test requires about 10 minutes to administer. It is suitable for a wide range of ages from 2 1/2 through adulthood and has established age norms based on a national sample of 2,725 children and adults tested at 240 sites across the U.S.

The PPVT-III has been extensively revised from earlier versions of the test. These improvements were undertaken to promote easier testing and more accurate scoring. Also, new drawings have been added and dated illustrations dropped so as to achieve better gender and ethnic balance. Individual test items that showed statistical bias by race or ethnicity, gender, or region were deleted from the item pool for the scale prior to standardization. PPVT-III was reported to be highly reliable utilizing FACES data with internal-consistency reliability (alpha) coefficients ranging from .96 for Fall, 2000 to .97 for Spring, 2001.

A Spanish-language test, the *Test de Vocabulario en Imagenes Peabody* (TVIP), is also available, but has not been updated to be directly comparable to the PPVT-III. For FACES, the TVIP was used with children whose primary language was Spanish.

A screener was used to determine whether English-language learners were to be administered the direct child assessment battery in English or not. The screener involved information provided by teachers and assessors which was used to determine the language of administration. In Fall 2000, English-language learners who were determined to be primarily Spanish-speaking, received the entire direct child assessment battery in Spanish, e.g. TVIP, Woodcock Munoz Letter-Word Identification, Applied Problems, Dictation, etc. They also were administered the PPVT and Woodcock Johnson Letter-Word Identification in English, as well. In Spring 2001, these same children received the entire direct child assessment battery in English. They were also administered the TVIP and Woodcock Munoz Letter Word Identification in Spanish for the purpose of comparison. In Fall 2000, English-language learners who were determined to primarily speak a language other than Spanish did not receive any portion of the direct child assessment battery in their native languages. In Spring 2001, these same children received the entire direct child assessment battery in English.

### A2. Woodcock-Johnson Psycho-Educational Battery – Revised

The updated edition of the Woodcock-Johnson Battery (WJ-R) is a carefully constructed and widely used test battery. The set of individually administered tests is designed to assess the intellectual and academic development of individuals from preschool through adulthood (Woodcock and Johnson, 1989; Salvia and Ysseldyke, 1991). FACES used three subtests from the Achievement Battery that together constitute an "Early Development -- Skills" cluster, according to the test developers. The cluster is comprised of the Letter-Word Identification, Applied Problems, and Dictation tests. The same three subtests of the Spanish version (*Woodcock-Muñox Pruebas de Aprovechamiento-Revisada*) were used in the Spanish version of the FACES assessment battery.

**Letter-Word Identification.** The first five Letter-Word Identification items involve symbolic learning, or the ability to match a rebus (pictographic representation of a word) with an actual picture of the object. The remaining items measure children's reading identification skills in identifying isolated letters and words that appear in large type on the pages of the test book. As well as being part of the Early Development cluster, this subtest is also part of the Basic Reading Skills cluster. The internal consistency of the Letter-Word Identification subtest with FACES children averaged .84 for Fall, 2000 and .86 for Spring, 2001.

**Letter Naming.** The Letter Naming task is a test developed for use in the Head Start Quality Research Centers curricular intervention studies. Children are shown all 26 upper-case letters of the alphabet, divided into three groups of 8, 9, and 9 letters, arranged in approximate order of item difficulty. They are asked to identify the letters they know by name. It has the virtue of providing specific numeric information about how many letters Head Start children learn and which ones they are more or less likely to acquire. The Letter Naming task provides complementary information to the Woodcock Johnson Letter Word Identification task regarding children's knowledge and awareness of letters. Children's knowledge and awareness of letters is an essential prerequisite to their learning how to read.

**Applied Problems.** This subtest measures children's skill in analyzing and solving practical problems in mathematics. In order to solve the problems, the child must recognize the procedure to be followed and then perform relatively simple counting, addition or subtraction operations. Because many of the problems include extraneous stimuli or information, the child must also decide which data to include in the count or calculation. As well as being part of the Early Development cluster, the subtest is also part of a Broad Mathematics cluster. The internal consistency of the Applied Problems subtest with FACES children averaged .90 for Fall, 2000 and .91 for Spring, 2001.

**Dictation.** The first six items in this subtest measure prewriting skills, such as drawing lines and copying letters. The remaining items measure the child's skill in providing written responses when asked to write specific upper- or lower-case letters of the alphabet. Later parts of the test ask the child to write specific words and phrases, punctuation, and capitalization. The internal consistency of the Dictation subtest with FACES children averaged .77 for both Fall, 2000 and Spring, 2001.

## A3. McCarthy Scales of Children's Abilities

The McCarthy Scales of Children's Abilities is a widely used and well-documented test battery. FACES employed one subtest from the battery, the Draw-A-Design Task. The Draw-A-Design Task was used to assess children's perceptual-motor skills. This task asks the child to draw copies of a series of increasingly complex geometric figures. For FACES, this task was directly translated as part of the Spanish version of the assessment. The Draw-A-Design Task was reported, utilizing FACES data, with internal-consistency reliability (alpha) coefficients ranging from .58 for Fall, 2000 to .68 for Spring, 2001.

## A4. Story and Print Concepts

The Story and Print Concepts task was an adaptation of earlier prereading assessment procedures developed by Marie Clay (1979), William Teale (1988, 1990), and Mason and Stewart (1989). In these procedures, a child is handed a children's storybook (FACES Battery - *Where's My Teddy?* (Alborough, 1992) or *¿Dónde Está Mi Osito?* (Alborough, Castro, Trans. 1992)) upside down and backwards. The assessor asks a series of questions designed to test the children's knowledge of books. These include questions regarding the location of the front of the book, the point at which one should begin reading, and information relating to the title and author of the book. The assessor reads the story to the child and asks basic questions about both the mechanics (print conventions) of reading and the content (comprehension) of the story. The print conventions questions pertain to children's knowledge of the left-to-right and up-and-down conventions of reading, while the comprehension questions pertain to children's recall of key facts from the story. Additionally, for FACES, questions were added tapping rhyming awareness (e.g., "I'll say some words from the story and you tell me whether they rhyme, OK - bawl and small, etc.") and phonological awareness (e.g., "What word would be left if I took "teh" away from Ted?"). These additions were only included in the Fall, 2000 direct child assessment battery. FACES reliabilities (internal consistencies) for these concepts for both the Fall and the Spring were as follows: 1.) Book Knowledge (.57 and .59); 2.) Print Conventions (.73 and .74); and 3.) Comprehension (.43 and .41).

### A5. Social Awareness

This measure was adapted from a subtest of the Comprehensive Assessment Program (CAP) Early Childhood Diagnostic Instrument used by Snow et al. (1995) among others to test children's general knowledge and awareness of the social environment. The child is asked to give his/her "full name," which includes both first and last name, his/her age (either verbally, which is given full credit or by holding up the correct number of fingers, which is given partial credit) and month/day of birth. The FACES reliabilities for the Social Awareness measure were .63 for Fall, 2000 and .61 for Spring, 2001.

### A6. Color Names and One-to-One Counting

This was also a subtest of the CAP Early Childhood Diagnostic Instrument used by Snow et al. (1995) and developed by Marie Clay (1979), William Teale (1988, 1990) and Mason and Stewart (1989) as a battery of emergent literacy and school readiness measures. For the FACES battery, 10 teddy bears of different colors are presented randomly arranged on a page and the child is asked to point to each in turn and name the color. Following the Color Names task, the child is asked to count the bears and the assessor marks the final number the child arrives at when finished counting (correct answer is "10"). After this, the child is asked to report the total number of bears. The verbatim response is then recorded. Following these questions, the assessor must rate the child's one-to-one counting performance using a 5-point scale. At the extremes, a score of 5 indicated that the child made no mistakes and score of 1 indicated that the child could not count or did not try to count. The FACES reliabilities for the Color Naming task were .95 for Fall, 2000 and .94 for Spring, 2001.

### A7. Leiter International Performance Scale - Revised (Leiter-R) - Attention Sustained

The Leiter-R by Roid and Miller (1997) assesses cognitive function in children and adolescents. The battery includes measures of nonverbal intelligence in fluid reasoning and visualization, as well as appraisals of visuospatial memory and attention. In Spring 2001, the Leiter-R AS (Attention Sustained) Subtest was added to the FACES direct child assessment battery to permit assessments of children's visuospatial memory and attention. The subtest is primarily nonverbal and is administered in two subsections – the first being for those 2-3 years of age and the second being for those 4-5 years of age. Assessors provide minimal instructions throughout the administration of the Leiter-R AS. Children are presented with a series of pages containing pictures and are instructed to mark off all pictures that resemble a reference picture. The assessor times the child, with times ranging from 30 seconds to 120 seconds allotted for completion of the tasks. FACES reliabilities for the Leiter Attention Sustained subtask by age groupings for Spring 2001 were as follows: 1.) 2 - 3 year old - .71; and 2.) 4 - 5 year olds - .81.

### A8. Interviewer Ratings

At the end of the one-on-one testing sessions with the children, the assessor completes a set of rating scales evaluating the child's behavior in the test situation, including the child's approaches to learning and problem behaviors. There are two sections to these ratings. The first consists of eight scales rating the child's response during the assessment on eight different domains: task persistence, attention span, body movement, attention to directions, comprehension of directions, verbalization, ease of relationship, and the child's level of confidence. Ratings use 4-point scales with descriptive anchors at each point. For example, the "task persistence" scale consists of the following anchor points: persists with task (4), attempts task briefly (3), attempts task after much encouragement (2), refuses (1). The FACES reliabilities for the Interviewer Ratings were .82 for Fall, 2000 and .81 for Spring, 2001.

The second section asks the assessor to indicate any special concerns regarding the child's ability to complete the assessment: responding nonverbally, using nonstandard English such as dialect, speaking English as a second language, having limited English proficiency, experiencing difficulty hearing or seeing the assessor/test materials, or reporting the child's speech was difficult to understand. These items use 3-point ratings to indicate the degree to which the child displayed any of these characteristics (i.e., "not at all," "somewhat," and "very much").

### A9. Kindergraten Follow-Up ELCS-K Measures

Two additional measures were included in the follow-up kindergarten assessment battery

(Spring, 2002): the Reading scale and the General Knowledge scale, which were adapted from the Early Childhood Longitudinal Study – Kindergarten Cohort (ECLS-K).

In ECLS-K, the Reading scale taps a variety of skills that indicate reading ability (including familiarity with print), recognition of letters and phonemes, vocabulary, and reading comprehension skills (e.g., children's understanding of the text), as well as their personal reflection and critical evaluation of the text. The General Knowledge scale taps skills in the natural sciences (e.g., their conceptual understanding of why things occur as they do, and their ability to pose questions and investigate answers in the natural sciences) and social studies (e.g., their basic knowledge of History, Government, and Culture). Both scales follow the guidelines of the 1996 National Assessment of Educational Progress, have been reviewed by curriculum experts, as well as elementary school teachers, and have been found to be both reliable and valid measures of reading achievement and basic knowledge acquisition.[29]

The Reading assessment was administered in two stages. First, a routing test was administered to estimate the child's reading ability. Based on his/her performance on the routing test (either "high," "medium," or "low"), an appropriate "second stage" test was administered. The Reading assessment had three levels of second stage tests: low (red), medium (yellow), and high (blue). For the General Knowledge assessment, each child was administered only the routing test. Estimates of reliability with FACES data, as measured by Cronbach's coefficient alpha, will be provided at a later point when the data become available.

---

[29] For more information on the Reading and General Knowledge measures, please refer to the Early Childhood Longitudinal Study website at http://nces.ed.gov/ecls/.

## B. Classroom Observation Instruments

In FACES, two distinctive types of observation instruments (i.e., classroom observation and child observation) were used to measure peer interactions, friendships of children, and the extent to which Head Start programs employed skilled teachers and provided developmentally appropriate environments and curricula for their children.

### B1. Counts of Children and Adults

The Counts of Children and Adults provide information needed to calculate child/adult ratios and for other calculations to be used in assessing specific measures of classroom quality. Classroom observers are tasked with counting the number of children (boys and girls), the number of paid staff, and the number of adult volunteers at two separate time periods during the classroom observation. The two counts must be at least an hour apart and must involve one structured (teacher-directed) activity and one unstructured activity.

### B2. Assessment Profile

The Assessment Profile (Abbott-Shim and Sibley, 1987) is a structured observation guide designed to provide a quantitative assessment of classrooms and teaching practices that facilitate the learning and development of children. Three subscales were used in FACES: Scheduling, Learning Environment, and Individualizing.

The **Scheduling** subscale assesses the written plans for classroom scheduling and how classroom activities are implemented. The appropriateness and completeness of the classroom activity plan are also noted. The subscale also assesses the balance and variety of learning contexts (e.g., individual, small group, and large group) and learning opportunities (i.e., child- vs. teacher-directed and active vs. quiet activities). The 14 observation items are scored in a yes/no format. High scores on this measure are indicative of a teacher that uses a "planful" approach to classroom activities. The reliability of the Scheduling subscale was reported as .89 for Fall, 2000 and .87 for Spring, 2001.

The **Learning Environment** subscale focuses on the accessibility of a variety of learning materials to children in the classroom. Variety is assessed across various conceptual areas, such as science, math, language, fine motor, etc. and also within each conceptual area. The subscale also assesses how classroom space is arranged to determine whether the classroom encourages independence (e.g., whether the learning materials are located on low shelves and clearly labeled) and reflects the child as an individual. When materials are both available and accessible, and in sufficient numbers (typically a minimum of three in each group) the item is given a positive score. High scores on this 7-item measure indicate a "learning rich" environment, filled with toys and learning materials that address a variety of developmental domains. The reliability of the Learning Environment subscale was reported as .68 for Fall, 2000 and .77 for Spring, 2001.

The **Individualizing** subscale focuses on the extent to which emphasis is placed on children, individually, in the classroom setting. This includes whether or not there are periodic individual assessments of each child's performance using portfolios of his/her work, performance inventories, and teacher notations. Also included, is whether or not child assessment information is used for planning individualized learning experiences. The final inclusion involves whether or not teachers have the ability to make provisions for children with special needs. The reliability of the Individualizing subscale was reported as .50 for Fall, 2000 and .54 for Spring, 2001.

### B3. Early Childhood Environment Rating Scale-Revised (ECERS-R)

The Early Childhood Environment Rating Scale-Revised (ECERS-R) is a global rating of classroom quality based on structural features of the classroom (Harms and Clifford, 1980). It has been widely used in child development research and has predicted optimal child outcomes in a number of studies (e.g., Phillips, Voran, Kisker, Howes and Whitebook, 1994). The revised version of the ECERS provides improvements to the items and represents an improvement on the standardization of the observational methods. In addition, the ECERS-R is easier to train and gain inter-rater reliability. The ECERS-R contains 37 items representative of classroom quality. Each item is coded on a 7-point scale with a score of 1 representing "inadequate", a score of 3 representing "minimal quality," a score of 5 representing "good quality," and a score of 7 representing "excellent quality." The internal consistency of the ECERS-R mean score for all combined items was .92 for both Fall, 2000 and Spring, 2001.

Seven subscales were derived from the ECERS-R for usage in analyses of FACES classroom quality, each pertaining to different elements of classroom quality. These are as follows: 1.) Personal Care Routines are measured using six items: greeting/departing, meals/snacks, nap/rest, toileting/diapering, health practices, and safety practices; 2.) Furnishings is measured using four items: indoor space, furniture for routine care, play, and learning, furniture for relaxation and comfort, and room arrangement for play; 3.) Language Skills are measured using four items: books and pictures, encouraging children to communicate, using language to develop reasoning skills, and informal use of language; 4.) Motor Skills are measured using four items: space for gross motor play, gross motor equipment, fine motor activities, and supervision of gross motor activities; 5.) Creativity is measured using six items: child-related display, art, music/movement, blocks, sand/water, and dramatic play; 6.) Social Skills are measured using four items: supervision, other than gross motor activity, discipline, staff-child interactions, and interactions among children; and 7.) Program Structure is measured using four items: space for privacy, schedule, free play, and group time. Five items were not incorporated into any of the subscales which are as follows: nature/science, math/numbers, use of TV, video, and/or computers, promoting acceptance of diversity, and provisions for children with disabilities. Thus there were only 32 of the 37 available items included in the subscales.

A separate subscale, labeled ECERS-R Language, was comprised of four items and was devised to assess the quality of the language environment in Head Start classrooms. Additional information about this subscale can be found in Chapter 4.

### B4. Classroom Observation of Teacher-Directed Activities

The Classroom Observation of Teacher-Directed Activities is a checklist completed by classroom observers of observed teacher-directed activities in 21 specific areas, e.g. reading stories, singing songs, etc. The classroom observer indicates whether observed activities were directed toward individual children (Individual Attention), a small group of children (Small Group = 3 to 8 children), or a whole group of children (Whole Group = entire classroom). Observers were instructed to mark down, only once for any item, any teacher-directed activities observed throughout the course of the classroom observation and if these observed activities were directed toward individuals, a small group of children, or the entire classroom. This checklist was introduced in Spring, 2001.

### B5. Arnett Caregiver Interaction Scale

The Arnett Caregiver Interaction Scale is a rating scale of teacher behavior towards the children in the classroom. It consists of 26 items that assess five areas of teacher behavior: sensitivity, punitiveness, detachment, permissiveness, and prosocial interaction (Arnett, 1989). The version of the Arnett Caregiver Interaction Scale utilized in the current round of FACES consists of 30 items and five subscales with the subscale labels being as follows: Sensitivity, Harshness, Detachment, Permissiveness, and Independence. At the end of the observational period, the observer completes the scale for an individual teacher, typically the lead teacher in the classroom. For example, in evaluating whether the teacher "speaks warmly to the children," the observer will assign ratings indicating the extent to which the statement is characteristic of the teacher, from 1 "never seen" to 4 "always or almost always." The Cronbach Coefficient Alpha for all of the items was .94 for Fall, 2000 and .69 for Spring, 2001.

## C. Teacher's Child Ratings and Teacher Background

Teacher ratings of children were important sources of information about children's learning and behavior because teachers see children over extended periods of time and in a variety of settings. Using a rating form known as the Teacher's Child Report (TCR), teacher's were first asked to rate each child on a set of behaviors that assessed the child's basic social skills and classroom behavior. In these two sections, the teacher is asked to indicate the extent to which a given statement (e.g., "follows the teacher's directions") is characteristic of the child, from 1 "never" to 3 "very often." The items making up these ratings form two scales:

### C1. Cooperative classroom behavior:

There are 12 ratings items for the teacher to indicate how often the child engages in cooperative classroom behaviors such as following teacher's directions, helping put things away, complimenting classmate, and following rules when playing games. The ratings include items drawn from the Personal Maturity Scale (Alexander and Entwisle, 1988) and the Social Skills Rating System (Elliott, Gresham, Freeman, and McCloskey, 1988) to assess positive behavior such as cooperation, sharing, and expression of feelings. A summary score is created from the 3-point scale items which ranges from zero to 24, with high scores indicating more frequent cooperative behavior. The internal consistency for this measure was .88 in both Fall, 2000 and Spring, 2001.

### C2. Total behavior problems:

The Behavior Problems scale is based on measures of negative child behaviors that are associated with learning problems and later grade retention. Items come from an abbreviated adaptation of the Personal Maturity Scale (Alexander and Entwisle, 1988), the Child Behavior Checklist for Preschool-Aged Children, Teacher Report (Achenbach, Edelbrock, and Howell, 1987) and The Behavior Problems Index (Zill, 1990). The items ask about the frequency of aggressive behavior (e.g., hits/fights with others), hyperactive behavior (e.g., is very restless), and anxious or depressed and withdrawn behavior (e.g., is unhappy). The summary score from the scale's 14 behavior items ranges from zero to 28, with higher scores representing more frequent or severe negative behavior. The reliabilities (internal consistency) for these measures for both Fall and Spring are as follows: 1.) Total Problem Behaviors - .86 for both; 2.) Aggression - .83 and .85; 3.) Hyperactivity - .72 for both; and 4.) .77 and .76.

The teacher is then asked to rate the child's problem solving skills and initiative, social relationships, creative representations, music/movement skills, and language/math skills. The teacher is asked to rate the child's highest level of behavior in each of the above domains observed in the past week. Scale points for each item are described on paper and there is a glossary that provides concrete examples of each anchor point. For the purpose of FACES, fourteen items from the Child Observation Record (COR; High/Scope Educational Research Foundation, 1992) were selected with a demonstrated reliability of .94 for both Fall, 2000 and Spring, 2001. These 14-items were further divided up into the following scales: social relationships, creative representations, music and movement, and cognitive.

### C3. **Social Relationships** (3 items):

A composite score based on teacher's ratings of how well the child makes friends, works with other children, and understands and expresses feelings. Each item is rated on a five-point scale with higher scores representing greater skill in coping with social situations and expressing feelings appropriately. The summary score is the average of the three items and ranges from one to five. The measure shows good reliability with the FACES study, with Alpha Coefficients of .83 for both Fall, 2000 and Spring, 2001.

### C4. **Creative Representations** (3 items):

A composite score based on the teacher's ratings of how well the child uses creative materials for self-expression in making and building things, drawing and painting, and engaging in pretend play. Each item is rated on a five-point scale with higher scores representing greater proficiency. The summary score is the average of the three items and ranges from one to five. The measure shows good reliability with the FACES study, with Alpha Coefficients of .80 for Fall, 2000 and .81 for Spring, 2001.

### C5. Music and Movement (4 items):

A composite score based on teacher's ratings of how well the child can imitate movements to a steady beat, follow music and movement directions, exhibit body coordination, and manipulate small objects and perform precise actions. Each item is rated on a five-point scale with higher scores representing greater proficiency. The summary score is the average of the four items and ranges from one to five. The measure shows good reliability with the FACES study, with Alpha Coefficients of .88 for both Fall, 2000 and Spring, 2001.

### C6. Cognitive (4 items):

A composite score based on teacher's ratings of how well the child can solve problems, engage in complex play, show interest in reading, and exhibit classification skills by sorting objects. Each item is rated on a five-point scale with higher scores representing greater proficiency. The summary score is the average of the four items and ranges from one to five. The measure shows good reliability with the FACES study, with Alpha Coefficients of .82 for Fall, 2000 and .83 for Spring, 2001.

The **Lead Teacher Background Information** consists of questions asking the teacher about himself/herself, including sociodemographic and educational background and professional experience. Information about the curriculum being used, his/her attitude and knowledge about early childhood education practice (see Teacher Beliefs Scale write-up referenced in Chapter 4), and accommodations he/she has made or that others have made to meet the learning needs of children in his/her classroom, particularly children with special needs are included, as well.

## D. Parent Interview

Data from the FACES Parent Interview, administered in Fall 2000 and Spring 2001, provide Head Start with a comprehensive understanding of the families that they serve, including the characteristics of households and household members, levels and types of participation in the program and in other community services, involvement with their children, and understanding of their children's development.

Parents were also asked to rate each child on a set of behaviors that assessed the child's basic social skills and behavior problems. In this section, the parent is asked to indicate the extent to which a given statement (e.g., "makes friends easily") is characteristic of the child, from 1 "not true" to 3 "very true or often true." The items making up these ratings were drawn from two well-known measures of children's positive behavior and behavior problems: the Entwisle scale of Personal Maturity (Entwisle, Alexander, Cadigan, and Pallis, 1987) and the Child Behavior Checklist for Preschool-Aged Children (Achenbach, Edelbrock, and Howell, 1987). Two scales were formed to assess children's social competence:

### D1. Social skills and positive approaches to learning:

Parents were asked to rate their child's social skills and positive approaches to learning by describing their children's skills in making friends and accepting their ideas, as well as enjoying learning and trying new things. A summary score based on the scale's seven items ranges from zero to 14, with higher scores representing more positive behavior. Table A-10 shows the reliabilities for the Social Skills measure in both Fall, 2000 and Spring, 2001.

### D2. Total Problem Behaviors:

Parents were also asked to rate their children on negative behaviors that are relatively common among preschool children and that are associated with adjustment problems in elementary school. Parents were asked about three domains of problem behavior: hyperactive behavior, aggressive behavior, and depressed or withdrawn behavior. The 12 behavior items were combined in a summary score ranging from zero to 24, with higher scores representing more frequent or severe negative behavior. Table A-10 shows the reliabilities for all of these behavior problem measures in both Fall, 2001 and Spring, 2001.

### D3. Other Parent Interview Scales/Measures Referenced in the Report:

**NAMES AND SOURCES FOR OTHER PARENT INTERVIEW SCALES/MEASURES REFERENCED IN THE REPORT**

| Name | Source |
|---|---|
| Pearlin Mastery Scale (Locus of Control) | Pearlin, L. I. and Schooler, C. (1978). The structure of coping. *Journal of Health and Social Behavior, 22,* 337-356. |
| CES-D Depression Scale | Radloff, L. S. (1977). The CES-D: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385-401. |
| Family Activities with Children | National Household Education Survey - FACES Research Team |
| Parental Involvement in Head Start | Head Start Quality Research Consortium (QRC) |
| Exposure to Violence | FACES Research Team |
| Domestic Violence Screener | Feldous, K. M., Koziol-McLain, J., Amsbury, H. L. et. al. (1997). Accuracy of three brief screening questions for detecting partner violence in the emergency room. *JAMA, 227(17),* 1357. |
| Substance Abuse Screener | Administration for Children and Families (1997). National Impact Evaluation of the Comprehensive Child Development Program. Washington, D.C.: U.S. Department of Health and Human Services. |
| Involvement with Criminal Justice System | FACES Research Team |
| Parenting Style | National Longitudinal Study of Youth (NLSY), Early Head Start Evaluation (EHS), QRC |

## V. Field Staff Training

A weeklong training was conducted prior to each data collection period to prepare field staff for successful completion of data collection. The training included a wide variety of activities covering all the procedures, techniques, and contents required to carry out successful data collection in the Head Start centers:

- Lecture, incorporating slides, overheads, and videotapes;
- Exercises that simulate various procedures such as assessing classroom scheduling;
- Video demonstration of assessment techniques and components of classroom scoring procedures;
- Exercises to achieve pre-established levels of inter-rater reliability;
- Participatory involvement of all trainees in small groups so that trainers may evaluate individual performance;
- Multiple occasions of practice in real classroom settings that simulate what they are expected to do in the field, with the presence of a trainer and a small group of trainees to discuss the classroom ratings and provide valuable guidance on scoring reliability and agreement; and
- One-on-one practice and role-play in the administration of child assessment procedures under supervision of training staff.

The field procedures manual contained information about working with a research team, appropriate behaviors within a classroom, and how to orchestrate Head Start center visits. Moreover, the manual covered an overview of all data collection instruments and administrative and travel procedures. Complete scoring rules and question-by-question specifications for the child assessment and child and classroom observation instruments were also discussed in the manual.

During the training, trainees were introduced to the purpose and goals of the study and background information on Head Start. Trainees were also introduced to the data collection materials and general issues regarding children and early childhood learning environments. Each day of training included a morning question and answer period regarding the previous day's training, a daily review of the current day's material, and a brief discussion of the next day's events.

An additional practice session was given to provide trainees with more practice in either observation or assessment. Assignment of this practice was based on the measures in which the trainees needed more practice. For administering child assessments in Spanish, a special training for English-Spanish speaking trainees was held. The bilingual trainees had an opportunity to practice assessments with Spanish-speaking children.

## VI. Data Collection Procedures

### A. Site Visit Arrangements

The research team obtained feasible dates for the 2-week site visit from each of the sampled Head Start programs. Site visit dates for each program were coordinated within the data collection period and programs were notified about the visit dates. Three weeks before the site visit, a scheduling packet which contained the final visit schedule, a master list, organized by classroom, a reminder list, and a request for maps and directions to aid the research team was sent to the on-site coordinator (OSC). OSC's are members of the Head Start program staff specially designated to coordinate the data collection efforts by scheduling parent interviews, classroom visits with program teachers, and obtaining consent forms, among other related duties.

## VII. Quality Control Visits

In FACES, Quality Control (QC) visits were built into every step of the data collection to ensure the highest quality data possible. The QC visitors consisted of the FACES project staff who were involved in designing the instruments, preparing the training materials, and conducting the training. The QC visitors were trained in both observation and assessment data collection and also served as technical consultants in the field. During the Fall 2000 data collection, one 3-day QC visit to program sites was made.

## VIII. Data Preparation & Data File Creation

### A. Data entry:

Key entry and verification were performed on the study instruments using a sophisticated production data entry system. This system provides entry form layout, application of edit specification, data verification control, and provides data entry quality and production reports.

### B. Frequency review:

The frequencies of responses to all data items (both individually and in conjunction with related data items) were reviewed to ensure that appropriate skip patterns were followed. Members of the data preparation team checked each item to make sure the correct number of responses was represented for all items. If a discrepancy was discovered, the problem case was identified and reviewed.

### C. Data edit:

To code and edit questionnaire data, an integrated collection of software was utilized. Through this system of software, coding manuals and codebooks were developed, data editing was performed, and SAS source code was generated.

### D. Data File Creation:

Data files were created and analyses performed to provide summaries and assessments of Head Start children and their families during this period and to assess the reliability and validity of information contained within the data collection instruments. Numerous derived variables were created to increase the magnitude and scope of analytical capabilities. The coding for these derived variables may be obtained upon request.

## IX. Reliability and Data Summary

In FACES, various data collection instruments were used to assess the accomplishments and behaviors of children in Head Start programs, as well as the educational and familial support that is provided to them. As noted in Section IV: Data Collection Instruments, these instruments are widely used and report mostly high reliabilities. The reliabilities for each data collection instrument and summaries for these data collection instruments are provided in the following Tables: Table A-2 – Tables A-11.

**Table A-1. Summary of Measures Administered from Fall, 2000 to Spring, 2001**

| **Fall, 2000 (Head Start)** | **Spring, 2001 (Head Start)** |
|---|---|
| Social Awareness | Social Awareness |
| PPVT-III / TVIP | PPVT-III / TVIP |
| McCarthy Draw-A-Design | McCarthy Draw-A-Design |
| ----- | Leiter-R AS (Attention Sustained) Subset |
| Color Names and Counting | Color Names and Counting |
| Woodcock Johnson (Munoz): Letter-Word Identification | Woodcock Johnson (Munoz): Letter-Word Identification |
| Woodcock Johnson (Munoz): Applied Problems | Woodcock Johnson (Munoz): Applied Problems |
| Woodcock Johnson (Munoz): Dictation | Woodcock Johnson (Munoz): Dictation |
| Story and Print Concepts | Story and Print Concepts |
| Interviewer Rating: Assessment Behavior | Interviewer Rating: Assessment Behavior |

## Table A-2. Reliability of Fall, 2000 and Spring, 2001 FACES Child Assessment Data – English Assessments Only (Spring, 2001 Leiter Results are for Children Assessed in Both Eng. & Span.)

| | Fall, 2000 | | | Spring, 2001 | | |
|---|---|---|---|---|---|---|
| Scales | Number of Items | Number of Cases | Cronbach Alphas | Number of Items | Number of Cases | Cronbach Alphas |
| Social Awareness | 5 | 2,068 | .63 | 5 | 1,948 | .61 |
| PPVT-III | 144 | 2,116 | .96 | 144 | 1,980 | .97 |
| McCarthy: Draw-A-Design | 9 | 2,068 | .58 | 9 | 1,943 | .68 |
| Leiter-R AS - Ages 2 to 3 | - | - | - | 4 | 406 | .71 |
| Leiter-R AS - Ages 4 to 5 | - | - | - | 4 | 1,758 | .81 |
| Color Names | 10 | 2,055 | .95 | 10 | 1,940 | .94 |
| WJR: Letter-Word Identification | 23 | 1,054 | .84 | 23 | 1,595 | .86 |
| WJR: Applied Problems | 23 | 1,054 | .90 | 23 | 1,595 | .91 |
| WJR Dictation | 12 | 1,054 | .77 | 12 | 1,595 | .77 |
| Story and Print Concepts: Print Conventions | 2 | 2,116 | .73 | 2 | 1,980 | .74 |
| Story and Print Concepts: Book Knowledge | 5 | 2,116 | .57 | 5 | 1,980 | .59 |
| Story and Print Concepts: Comprehension | 2 | 2,116 | .43 | 2 | 1,980 | .41 |
| Interviewer Rating: Assessment Behavior | 8 | 2,021 | .82 | 8 | 1,901 | .81 |

**Table A-3. Reliability of Fall, 2000 and Spring, 2001 FACES Child Assessment Data – Spanish Assessments Only**
**(Spring, 2001 Leiter Results are Referenced in Table A-2.)**

| | Fall, 2000 | | | Spring, 2001 | | |
|---|---|---|---|---|---|---|
| Scales | Number of Items | Number of Cases | Cronbach Alphas | Number of Items | Number of Cases | Cronbach Alphas |
| Social Awareness | 5 | 385 | .36 | 5 | 356 | .45 |
| TVIP | 144 | 392 | .92 | 144 | 364 | .92 |
| McCarthy: Draw-A-Design | 9 | 375 | .57 | 9 | 355 | .74 |
| Leiter-R AS - Ages 2 to 3 | - | - | - | - | - | - |
| Leiter-R AS - Ages 4 to 5 | - | - | - | - | - | - |
| Color Names | 10 | 378 | .92 | 10 | 358 | .93 |
| WM: Letter-Word Identification | 23 | 219 | .75 | 23 | 307 | .78 |
| WM: Applied Problems | 23 | 219 | .85 | 23[1] | 307 | .89 |
| WM: Dictation | 12 | 219 | .77 | 12[1] | 307 | .73 |
| Story and Print Concepts: Print Conventions | 2 | 392 | .59 | 2 | 364 | .77 |
| Story and Print Concepts: Book Knowledge | 5 | 392 | .43 | 5 | 364 | .48 |
| Story and Print Concepts: Comprehension | 2 | 392 | .39 | 2 | 364 | .43 |
| Interviewer Rating: Assessment Behavior | 8 | 372 | .77 | 8 | 353 | .68 |

[1] Spring 2001 Applied Problems & Dictation are Woodcock Johnson, not Woodcock Munoz

### Table A-4. Summary Statistics for Fall, 2000 and Spring, 2001 FACES Child Assessment Data – English Assessments Only (Spring, 2001 Leiter Results are for Children Assessed in Both Eng. & Span.)

| | Fall, 2000 | | | | | Spring, 2001 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Scales | Number of Cases | Mean | SD | Reported Response Range | Possible Response Range | Number of Cases | Mean | SD | Reported Response Range | Possible Response Range |
| Social Awareness | 2,101 | 3.36 | 1.69 | 0 - 6 | 0 - 6 | 1,967 | 3.98 | 1.58 | 0 - 6 | 0 - 6 |
| PPVT-III* | 2,031 | 35.06 | 17.65 | 0 - 98 | 0 - 144 | 1,932 | 45.30 | 18.72 | 1 - 98 | 0 - 144 |
| McCarthy: Draw-A-Design | 2,112 | 2.92 | 1.33 | 0 - 13 | 0 - 19 | 1,980 | 3.52 | 1.70 | 0 - 15 | 0 - 19 |
| Leiter-R AS - Ages 2 to 5 | - | - | - | - | - | 2,253 | 40.72 | 10.79 | 1 - 70 | 0 - 70 |
| Color Names | 2,101 | 11.32 | 7.37 | 0 - 20 | 0 - 20 | 1,969 | 15.59 | 5.98 | 0 - 20 | 0 - 20 |
| WJR: Letter-Word Identification* | 948 | 5.30 | 2.61 | 0 - 21 | 0 - 23 | 1,511 | 6.59 | 3.19 | 0 - 22 | 0 - 23 |
| WJR: Applied Problems* | 963 | 7.52 | 4.36 | 0 - 21 | 0 - 23 | 1,542 | 8.98 | 4.70 | 0 - 22 | 0 - 23 |
| WJR: Dictation* | 916 | 5.11 | 1.83 | 0 - 12 | 0 - 12 | 1,491 | 5.64 | 2.11 | 0 -12 | 0 - 12 |
| Story and Print Concepts: Print Conventions | 2,089 | 0.23 | 0.57 | 0 - 2 | 0 - 2 | 1,968 | 0.37 | 0.69 | 0 - 2 | 0 - 2 |
| Story and Print Concepts: Book Knowledge | 2,087 | 1.62 | 1.27 | 0 - 5 | 0 - 5 | 1,961 | 2.41 | 1.30 | 0 -5 | 0 - 5 |
| Story and Print Concepts: Comprehension | 2,102 | 0.54 | 0.70 | 0 - 2 | 0 - 2 | 1,967 | 0.71 | 0.75 | 0 - 2 | 0 - 2 |
| Interviewer Rating: Assessment Behavior | 2,094 | 17.14 | 5.02 | 0 -24 | 0 - 24 | 1,950 | 19.05 | 4.38 | 0 - 24 | 0 - 24 |

*Raw scores were used.

**Table A-5. Summary Statistics for Fall, 2000 and Spring, 2001 FACES Child Assessment Data Spanish Assessments Only (Spring, 2001 Leiter Results are Referenced in Table A-4.)**

| | Fall, 2000 | | | | | Spring, 2001 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Scales | Number of Cases | Mean | SD | Reported Response Range | Possible Response Range | Number of Cases | Mean | SD | Reported Response Range | Possible Response Range |
| Social Awareness | 390 | 2.62 | 1.21 | 0 - 6 | 0 - 6 | 360 | 2.56 | 1.30 | 0 - 6 | 0 - 6 |
| TVIP* | 369 | 11.34 | 8.38 | 1 - 47 | 0 - 144 | 322 | 16.27 | 10.04 | 1 - 48 | 0 - 144 |
| McCarthy: Draw-A-Design | 392 | 3.37 | 1.34 | 0 - 13 | 0 - 19 | 364 | 4.05 | 1.90 | 0 -12 | 0 - 19 |
| Leiter-R AS - Ages 2 to 5 | - | - | - | - | - | - | - | - | - | - |
| Color Names | 386 | 8.90 | 6.62 | 0 - 20 | 0 - 20 | 362 | 13.46 | 6.49 | 0 - 20 | 0 - 20 |
| WM: Letter-Word Identification* | 195 | 4.37 | 1.20 | 0 - 10 | 0 - 23 | 295 | 5.01 | 1.68 | 0 - 12 | 0 - 23 |
| WM: Applied Problems* | 200 | 5.29 | 3.40 | 0 - 14 | 0 - 23[1] | 294 | 5.81 | 4.13 | 0 - 19 | 0 - 23 |
| WM: Dictation* | 188 | 4.99 | 1.28 | 1 - 11 | 0 - 12[1] | 293 | 5.72 | 1.74 | 0 - 11 | 0 - 12 |
| Story and Print Concepts: Print Conventions | 391 | 0.17 | 0.53 | 0 - 3 | 0 - 2 | 360 | 0.14 | 0.47 | 0 - 2 | 0 - 2 |
| Story and Print Concepts: Book Knowledge | 376 | 1.25 | 1.13 | 0 - 5 | 0 - 5 | 355 | 1.70 | 1.13 | 0 - 5 | 0 - 5 |
| Story and Print Concepts: Comprehension | 386 | 0.47 | 0.67 | 0 -2 | 0 - 2 | 360 | 0.50 | 0.68 | 0 - 2 | 0 - 2 |
| Interviewer Rating: Assessment Behavior | 383 | 17.54 | 4.41 | 0 -24 | 0 - 24 | 360 | 17.99 | 3.50 | 3 - 24 | 0 - 24 |

*Raw scores were used.

[1] Spring 2001 Applied Problems & Dictation are Woodcock Johnson, not Woodcock Munoz

**Table A-6. Reliability of Fall, 2000 and Spring, 2001 FACES Classroom Observation Data Selected Measures**

| | Fall, 2000 | | | Spring, 2001 | | |
|---|---|---|---|---|---|---|
| Scales | Number of Items | Number of Cases | Cronbach Alphas | Number of Items | Number of Cases | Cronbach Alphas |
| Assessment Profile: Scheduling | 14 | 227 | .89 | 14 | 243 | .87 |
| Assessment Profile: Learning Environment | 18 | 228 | .68 | 18 | 228 | .77 |
| Assessment Profile: Individualizing | 5 | 250 | .50 | 5 | 250 | .54 |
| ECERS Total Mean | 37 | 270 | .92 | 37 | 235 | .92 |
| Personal Care | 6 | 146 | .73 | 6 | 269 | .70 |
| Furnishings | 4 | 263 | .52 | 4 | 263 | .60 |
| Language | 4 | 260 | .77 | 4 | 272 | .76 |
| Motor Skills | 4 | 248 | .67 | 4 | 274 | .64 |
| Creative | 6 | 253 | .60 | 6 | 262 | .71 |
| Social | 4 | 264 | .86 | 4 | 269 | .91 |
| Program Structure | 4 | 256 | .60 | 4 | 261 | .69 |
| Arnett Scale of Caregiver Behavior: Lead Teacher (Total) | 30 | 256 | .94 | 30 | 258 | .69 |
| Sensitivity | 10 | 262 | .94 | 10 | 266 | .93 |
| Harshness | 9 | 265 | .83 | 9 | 271 | .66 |
| Detachment | 4 | 266 | .71 | 4 | 272 | .72 |
| Permissiveness | 3 | 266 | .52 | 3 | 274 | .20 |
| Independence | 4 | 263 | .58 | 4 | 269 | .24 |

**Table A-7. Summary Statistics for Fall, 2000 and Spring, 2001 FACES Classroom Observation Data Selected Measures**

| | Fall, 2000 | | | | | Spring, 2001 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Scales | Number of Cases | Mean | SD | Reported Response Range | Possible Response Range | Number of Cases | Mean | SD | Reported Response Range | Possible Response Range |
| Assessment Profile: Scheduling | 270 | 11.12 | 3.20 | 0 -14 | 0 - 14 | 270 | 11.00 | 3.15 | 3 - 14 | 0 - 14 |
| Assessment Profile: Learning Environment | 270 | 14.44 | 2.57 | 5 - 18 | 0 - 18 | 274 | 14.22 | 2.92 | 4 - 18 | 0 - 18 |
| Assessment Profile: Individualizing | 260 | 3.58 | 1.15 | 0 - 5 | 0 - 5 | 271 | 3.42 | 1.11 | 0 - 5 | 0 – 5 |
| ECERS Total Mean | 258 | 4.84 | 0.87 | 1.8 – 6.7 | 1.0 - 7.0 | 265 | 4.91 | 1.00 | 1.9 - 6.9 | 1.0 - 7.0 |
| Personal Care | 261 | 5.22 | 1.33 | 1.0 - 7.0 | 1.0 - 7.0 | 273 | 4.84 | 1.63 | 1.0 - 7.0 | 1.0 -7.0 |
| Furnishings | 268 | 5.48 | 1.01 | 1.3 - 7.0 | 1.0- 7.0 | 266 | 5.52 | 1.02 | 2.3 - 7.0 | 1.0 -7.0 |
| Language | 270 | 4.86 | 1.20 | 1.0 - 7.0 | 1.0 -7.0 | 274 | 5.01 | 1.27 | 1.0 - 7.0 | 1.0 -7.0 |
| Motor Skills | 264 | 4.79 | 1.27 | 1.3 - 7.0 | 1.0 -7.0 | 277 | 4.89 | 1.34 | 1.3 - 7.0 | 1.0 -7.0 |
| Creative | 268 | 4.31 | 0.90 | 2.0 - 6.3 | 1.0 -7.0 | 277 | 4.26 | 1.02 | 1.8 - 6.7 | 1.0 -7.0 |
| Social | 266 | 5.39 | 1.38 | 1.3 - 7.0 | 1.0 -7.0 | 276 | 5.59 | 1.56 | 1.0 - 7.0 | 1.0 -7.0 |
| Program Structure | 268 | 4.87 | 1.24 | 1.0 - 7.0 | 1.0 -7.0 | 276 | 5.12 | 1.28 | 1.0 - 7.0 | 1.0 -7.0 |
| Arnett Scale of Caregiver Behavior: Lead Teacher (Total) | 267 | 71.48 | 12.42 | 26 - 90 | 0 - 90 | 274 | 72.84 | 12.99 | 20 - 89 | 0 - 90 |
| Sensitivity | 262 | 20.99 | 6.52 | 1 - 30 | 0 - 30 | 266 | 21.74 | 6.42 | 3 - 30 | 0 - 30 |
| Harshness | 265 | 24.28 | 3.22 | 4 - 27 | 0 - 27 | 271 | 24.34 | 3.16 | 6 - 27 | 0 - 27 |
| Detachment | 266 | 11.06 | 1.54 | 4 - 12 | 0 - 12 | 272 | 11.08 | 1.79 | 2 - 12 | 0 - 12 |
| Permissiveness | 266 | 7.47 | 1.37 | 1 - 9 | 0 -9 | 274 | 7.39 | 1.45 | 2 - 9 | 0 - 9 |
| Independence | 263 | 7.77 | 2.31 | 1 - 12 | 0 -12 | 269 | 8.39 | 2.32 | 1 - 12 | 0 - 12 |

**Table A-8. Reliability of Fall, 2000 and Spring, 2001 FACES Teacher's Child Report Data Selected Measures**

| | Fall, 2000 | | | Spring, 2001 | | |
|---|---|---|---|---|---|---|
| Scales | Number of Items | Number of Cases | Cronbach Alphas | Number of Items | Number of Cases | Cronbach Alphas |
| Social Skills | 12 | 2,522 | .88 | 12 | 2,254 | .88 |
| Behavioral Problems (Total) | 14 | 2,522 | .86 | 14 | 2,254 | .86 |
| Withdrawn | 7 | 2,522 | .77 | 7 | 2,254 | .76 |
| Aggressive | 4 | 2,522 | .83 | 4 | 2,254 | .85 |
| Hyperactive | 3 | 2,522 | .72 | 3 | 2,254 | .72 |
| Child Observation Record (Total) | 14 | 2,522 | .94 | 14 | 2,254 | .94 |
| Social Relationships | 3 | 2,522 | .83 | 3 | 2,254 | .83 |
| Creative Representations | 3 | 2,522 | .80 | 3 | 2,254 | .81 |
| Music and Movement | 4 | 2,522 | .88 | 4 | 2,254 | .88 |
| Cognitive | 4 | 2,522 | .82 | 4 | 2,254 | .83 |

## Table A-9. Summary Statistics for Fall, 2000 and Spring, 2001 FACES Teacher's Child Report Data Selected Measures

| | Fall, 2000 | | | | | Spring, 2001 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Scales | Number of Cases | Mean | SD | Reported Response Range | Possible Response Range | Number of Cases | Mean | SD | Reported Response Range | Possible Response Range |
| Social Skills | 2,518 | 14.52 | 4.85 | 0 – 24 | 0 - 24 | 2,251 | 16.58 | 4.63 | 0 -24 | 0 - 24 |
| Behavioral Problems (Total) | 2,484 | 5.76 | 5.06 | 0 – 28 | 0 - 28 | 2,215 | 5.30 | 4.89 | 0 - 27 | 0 - 28 |
| Withdrawn | 2,453 | 2.59 | 2.65 | 0 – 14 | 0 - 14 | 2,186 | 2.34 | 2.49 | 0 -14 | 0 - 14 |
| Aggressive | 2,449 | 1.80 | 2.04 | 0 – 8 | 0 - 8 | 2,189 | 1.74 | 2.03 | 0 - 8 | 0 - 8 |
| Hyperactive | 2,466 | 1.37 | 1.53 | 0 – 6 | 0 - 6 | 2,194 | 1.21 | 1.47 | 0 - 6 | 0 - 6 |
| Child Observation Record (Total) | 2,403 | 2.82 | 0.77 | 1 – 5 | 1 - 5 | 2,174 | 3.50 | 0.81 | 1 - 5 | 1 - 5 |
| Social Relationships | 2,464 | 2.83 | 1.01 | 1 – 5 | 1 - 5 | 2,230 | 3.51 | 0.99 | 1 - 5 | 1 - 5 |
| Creative Representations | 2,459 | 2.83 | 0.83 | 1 – 5 | 1 - 5 | 2,230 | 3.52 | 0.87 | 1 - 5 | 1 - 5 |
| Music and Movement | 2,446 | 2.95 | 0.85 | 1 – 5 | 1 - 5 | 2,217 | 3.65 | 0.89 | 1 - 5 | 1 - 5 |
| Cognitive | 2,403 | 2.75 | 0.82 | 1 – 5 | 1 - 5 | 2,184 | 3.41 | 0.89 | 1 - 5 | 1 - 5 |

**Table A-10. Means and Standard Deviations for Scales From the Parent Interview and the Teachers' Ratings of Children's Behavior**

| Scales | Fall 2000 | | | Spring 2001 | | |
|---|---|---|---|---|---|---|
| | Mean[1] | SD | N | Mean | SD | N |
| CES-D | 6.76 | 6.67 | 2473 | 6.57 | 6.31 | 2285 |
| Pearlin-Mastery Score | 14.80 | 3.27 | 2471 | 15.28 | 3.31 | 2284 |
| Exposure to Violence | 6.08 | 1.85 | 2485 | Not Asked | | |
| Weekly Activities with Children | 5.91 | 1.14 | 2485 | 6.10 | 1.05 | 2290 |
| Monthly Activities with Children | 1.93 | 1.48 | 2485 | 2.30 | 1.58 | 2290 |
| Emergent Literacy | 1.97 | 1.45 | 2469 | 2.99 | 1.50 | 2290 |
| Authoritative Parenting Style | 4.22 | 0.61 | 2483 | 4.23 | 0.59 | 2284 |
| Authoritarian Parenting Style | 2.20 | 0.71 | 2482 | 2.17 | 0.69 | 2283 |
| | | | | | | |
| Positive Social Behavior – Parent Report | 12.14 | 1.74 | 2479 | 12.14 | 1.77 | 2287 |
| Problem Behavior– Parent Report | 6.16 | 3.57 | 2480 | 5.61 | 3.56 | 2287 |
| Aggressive Behavior– Parent Report | 3.15 | 1.74 | 2468 | 2.83 | 1.73 | 2275 |
| Hyperactive Behavior– Parent Report | 1.86 | 1.51 | 2471 | 1.65 | 1.47 | 2275 |
| Withdrawn Behavior– Parent Report | 0.60 | 0.94 | 2459 | 0.62 | 0.93 | 2267 |
| | | | | | | |
| Problem Behavior– Teacher Report | 5.64 | 5.01 | 2484 | 5.23 | 4.89 | 2215 |
| Aggressive Behavior– Teacher Report | 1.75 | 2.02 | 2449 | 1.72 | 2.03 | 2189 |
| Hyperactive Behavior– Teacher Report | 1.36 | 1.52 | 2466 | 1.20 | 1.46 | 2194 |
| Withdrawn Behavior– Teacher Report | 2.51 | 2.64 | 2453 | 2.30 | 2.48 | 2186 |
| | | | | | | |
| Satisfaction with Head Start | --- | --- | --- | 30.88 | 4.15 | 2185 |
| Positive Experience with Head Start | --- | --- | --- | 49.97 | 6.71 | 2185 |
| Involvement with Head Start | --- | --- | --- | 31.30 | 6.02 | 2181 |

[1] Means and Standard Deviations are based on weighted data.

**References:**

Abbott-Shim, M. and Sibley, A. (1987). *Assessment profile for early childhood programs.* Atlanta, Georgia: Quality Assist Inc.

Achenbach, T. M., Edelbrock, C. and Howell, C. T. (1987). Empirically based assessment of the behavioral/emotional problems of 2-3 year-old children. *Journal of Abnormal Child Psychology*, 15, 629-650.

Alborough, J. (1992). *Where's My Teddy?.* London, England: Walker Books Ltd.

Alborough, J. and Castro, M. (1995). *?Donde Esta Mi Osito?.* Miami, Fl: Santillana USA Publishing Company, Inc.

Alexander, K. L. and Entwisle, D. R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53 (2),* Serial No. 218.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-552.

Clay, M. M. (1979). *What did I write?* Auckland, N.Z.; Exeter, NH: Heinemann.

Dunn, L. M., & Dunn. L. M. (1997). *Peabody Picture Vocabulary Test, Third Edition. Examiner's manual and norms booklet.* Circle Pines, MN: American Guidance Service.

Elliot, S. N., Gresham, F. M., Freeman, R. and McCloskey, G. (1988). Teacher and observer ratings of children's social skills: Validation of the social skills rating scales. *Journal of Psychoeducational Assessment, 6,* 152-161.

Entwistle, D. R., Alexander, K. L., Cadigan, D., and Pallis, P. M. (1987). The emergent academic self-image of first graders: Its response to social structure. *Child Development*, 58, 1190-1206.

Harms, T., and Clifford, R. (1980). *Early childhood environment rating scale.* New York: Teachers College Press.

High/Scope Educational Research Foundation. (1992). *Child Observation Record-Manual.* Ypsilanti, MI: High/Scope Educational Research Foundation.

Mason, J. and Stewart, J. (1989). *Early childhood diagnostic instrument.* Iowa City, IA: American Testronics.

Phillips, D. A., Voran, M., Kisker, E., Howes, C., and Whitebook, M. (1994). Child care for children in poverty: Opportunity or inequity? *Child Development*, 65, 472-492.

Roid, G. H. and Miller, L. J. (1997). Examiners Manual: Leiter International Performance Scale

- Revised. Wood Dale, Il: Stoelting Company

Salvia, J. and Ysseldyke, J. E. (1991). *Assessment in Special and Remedial Education*, 5th ed. Boston: Houghton Mifflin Co.

Snow, C. E., Tabors, P. O., Nicholson, P. A., & Kurland, B. F. (1995). SHELL: Oral language and early literacy skills in kindergarten and first-grade children. *Journal of Research in Early Childhood Education, 10*, 37-48.

Teale, W. H. (1988). Developmentally appropriate assessment of reading and writing in early childhood classroom. *Elementary School Journal*, 89, 173-183.

Teale, W. (1990). Issues in early childhood assessment. In P.Afflerbach (Ed.). *Issues in statewide reading assessment*, 35-53. Washington, DC: ERIC/American Institutes for Research.

Woodcock, R W. and Johnson, M. B. (1989). *Woodcock-Johnson Revised Tests of Achievement*. Itasca, IL: Riverside Publishing.

Zill, N. (1990). Behavior Problems Index based on parent report [memorandum]. Washington, D.C.: Child Trends.

**U.S. Department of Education**
*Office of Educational Research and Improvement (OERI)*
*National Library of Education (NLE)*
*Educational Resources Information Center (ERIC)*

ERIC
Educational Resources Information Center

# NOTICE

## Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (1/2003)
PS031432